PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
A61K 39/12, C07H 21/04
C07K 7/06, 7/08, 7/10
C07K 13/00, 15/28

A1

(11) International Publication Number: **WO 90/11089**

(43) International Publication Date: 4 October 1990 (04.10.90)

(21) International Application Number: PCT/US90/01348

(22) International Filing Date: 15 March 1990 (15.03.90)

(30) Priority data:

325,338	17 March 1989 (17.03.89)	US
341,334	20 April 1989 (20.04.89)	US
355,002	18 May 1989 (18.05.89)	US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: HOUGHTON, Michael ; 53 Rosemead Court, Danville, CA 94526 (US). CHOO, Qui-Lim ; 5700 Fern Street, El Cerrito, CA 94530 (US). KUO, George ; 1370 Sixth Avenue, San Francisco, CA 94122 (US).

(74) Agents: MONROY, Gladys, H. et al.; Irell & Manella, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AU, FI, HU, JP, KR, NO, SU.

Published
With international search report.

(54) Title: NANBV DIAGNOSTICS AND VACCINES

(57) Abstract

A new virus, Hepatitis C virus (HCV), which has proven to be the major etiologic agent of blood-borne NANBH, was discovered by Applicant. The initial work on this virus, which includes a partial genomic sequence of the prototype HCV isolate, is described in EPO Pub. No. 318,216, and PCT Pub. No. WO 89/04669. The present invention, which in part is based on new HCV sequences and polypeptides which are not disclosed in the above-cited publications, includes the application of these new sequences and polypeptides in immunoassays, probe diagnostics, anti-HCV antibody production, PCR technology, and recombinant DNA technology. Included within the invention also are novel, immunogenic polypeptides encoded within clones containing HCV cDNA, novel methods for purifying an immunogenic HCV polypeptide, and antisense polynucleotides derived from HCV cDNA.

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CM	Cameroon
DE	Germany, Federal Republic of
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
HU	Hungary
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
RO	Romania
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

NANBV DIAGNOSTICS AND VACCINES

Technical Field

The invention relates to materials and methodologies for managing the spread of non-A, non-B hepatitis virus (NANBV) infection. More specifically, it relates to polynucleotides derived from the genome of an etiologic agent of NANBH, hepatitis C virus (HCV), to polypeptides encoded therein, and to antibodies directed to the polypeptides. These reagents are useful as screening agents for HCV and its infection, and as protective agents against the disease.

References Cited in the Application

Barr et al. (1986), Biotechniques 4:428.

Botstein (1979), Gene 8:17.

Brinton, M.A. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.327-374.

Broach (1981) in: Molecular Biology of the Yeast Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.

Broach et al. (1983), Meth. Enz. 101:307.

Chang et al. (1977), Nature 198:1056.

Chirgwin et al. (1979), Biochemistry 18:5294.

Chomczynski and Sacchi (1987), Analytical Biochemistry 162:156.

Clewell et al. (1969), Proc. Natl. Acad. Sci. USA 62:1159.

Clewell (1972), J. Bacteriol. 110:667.
Cohen (1972), Proc. Natl. Acad. Sci. USA 69:2110.
Cousens et al. (1987), gene 61:265.
De Boer et al. (1983), Proc. Natl. Acad. Sci. USA 292:128.
Dreesman et al. (1985), J. Infect. Disease 151:761.
Feinstone, S.M. and Hoofnagle, J.H. (1984), New Engl. J. Med. 311:185.
Fields & Knipe (1986), FUNDAMENTAL VIROLOGY (Raven Press, N.Y.).
Fiers et al. (1978), Nature 273:113.
Gerety, R.J. et al., in VIRAL HEPATITIS AND LIVER DISEASE (Vyas, B.N., Dienstag, J.L., and Hoofnagle, J.H., eds, Grune and Stratton, Inc., 1984) pp 23-47.
Goeddel et al. (1980), Nucleic Acids Res. 8:4057.
Graham and Van der Eb (1978), Virology 52:546.
Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961.
Grych et al. (1985), Nature 316:74.
Gubler and Hoffman (1983), Gene 25:263.
Hahn (1988) Virology 162:167.
Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS.
Han (1987), Biochemistry 26:1617.
Helfman (1983), Proc. Natl. Acad. Sci. USA 80:31.
Hess et al. (1968), J. Adv. Enzyme Reg 7:149.
Hinnen et al. (1978), Proc. Natl. Acad. Sci. 75:1929.
Hitzeman et al. (1980), J. Biol. Chem. 255:2073.
Holland et al. (1978), Biochemistry 17:4900.
Holland (1981), J. Biol. Chem. 256: 1385.
Houghton et al. (1981), Nucleic Acids Res. 9:247
Hunyh, T.V. et al. (1985) in DNA CLONING TECHNIQUES; A PRACTICAL APPROACH (D. Glover, Ed., IRL Press, Oxford, U.K.) pp. 49-78.
Immun. Rev. (1982) 62:185.
Iwarson (1987), British Medical J. 295:946.

Kennett et al. (1980) MONOCLONAL ANTIBODIES.
Kyte and Doolittle (1982), J. Mol. Biol. 157:105-132.
Laemmli (1970), Nature 227, 680.
Lee et al. (1988), Science 239:1288.
Maniatis, T., et al. (1982) MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).
Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London).
Maxam et al. (1980), Methods in Enzymology 65:499.
MacNamara et al. (1984), Science 226:1325.
Messing et al. (1981), Nucleic Acids Res. 9:309.
Messing (1983), Methods in Enzymology 101:20-37.
METHODS IN ENZYMOLOGY (Academic Press).
Michelle et al., Int. Symposium on Viral Hepatitis.
Monath (1986) in THE VIRUSES: THE TOGAVIRADAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.375-440.
Nagahuma et al. (1984), Anal. Biochem. 141:74.
Neurath et al. (1984), Science 224:392.
Nisonoff et al. (1981), Clin. Immunol. Immunopathol. 21:397-406.
Overby, L.R. (1985), Curr. Hepatol. 5:49.
Overby, L.R. (1986), Curr. Hepatol. 6:65.
Overby, L.R. (1987), Curr. Hepatol. 7:35.
Peleg (1969), Nature 221:193.
Pfefferkorn and Shapiro (1974), in COMPREHENSIVE VIROLOGY, Vol. 2 (Fraenkel-Conrat & Wagner, eds., Plenum, N.Y.) pp. 171-230.
Prince, A.M. (1983), Annu. Rev. Microbiol. 37:217.
Rice et al. (1985), Science 229:726.

Rice et al. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.279-328.

Roehrig (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press)

Sadler et al. (1980), Gene 8, 279.

Saiki et al. (1986), Nature 324: 163.

Saiki et al. (1988), Science 239:487.

Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463.

Schlesinger et al. (1986), J. Virol. 60:1153.

Schreier, M., et al. (1980) HYBRIDOMA TECHNIQUES

Scopes (1984), PROTEIN PURIFICATION, PRINCIPLES AND PRACTICE, SECOND EDITION (Springer-Verlag, N.Y.).

Shimatake et al. (1981), Nature 292:128.

Sippel (1973), Eur. J. Biochem. 37:31.

Steimer et al. (1986), J. Virol. 58:9.

Stollar (1980), in THE TOGAVIRUSES (R.W. Schlesinger, ed., Academic Press, N.Y.), pp. 584-622.

Sumiyoshi et al. (1987), Virology 161:497.

Taylor et al. (1976), Biochem. Biophys. Acta 442:324.

Towbin et al. (1979), Proc. Natl. Acad. Sci. USA 76, 4350.

Tsu and Herzenberg (1980), in SELECTED METHODS IN CELLULAR IMMUNOLOGY (W.H. Freeman and Co.) pp. 373-391.

Vytdehaag et al. (1985), J. Immunol. 134:1225.

Valenzuela, P., et al. (1982), Nature 298:344.

Valenzuela, P., et al. (1984), in HEPATITIS B (Millman, I., et al., ed, Plenum Press) pp. 225-236.

Warner (1984), DNA 3:401.

Wu and Grossman (1987), Methods in Enzymology Vol. 154, RECOMBINANT DNA, Part E.

Wu (1987), Methods in Enzymology vol 155, RECOMBINANT DNA, part F.

Zoller (1982), Nucleic Acids Res. 10:6487.

Cited Patents

EPO Pub. No. 318,216
PCT Pub. No. WO 89/04669
U.S. Patent No. 4,341,761
U.S. Patent No. 4,399,121
U.S. Patent No. 4,427,783
U.S. Patent No. 4,444,887
U.S. Patent No. 4,466,917
U.S. Patent No. 4,472,500
U.S. Patent No. 4,491,632
U.S. Patent No. 4,493,890

Background Art

Non-A, Non-B hepatitis (NANBH) is a transmissible disease or family of diseases that are believed to be viral-induced, and that are distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). NANBH was first identified in transfused individuals. Transmission from man to chimpanzee and serial passage in chimpanzees provided evidence that NANBH is due to a transmissible infectious agent or agents.

Epidemiologic evidence is suggestive that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of agents which may be the causative of NANBH are unknown.

Clinical diagnosis and identification of NANBH has been accomplished primarily by exclusion of other viral markers. Among the methods used to detect putative NANBV antigens and antibodies are agar-gel diffusion,

counterimmunoelectrophoresis, immunofluorescence microscopy, immune electron microscopy, radioimmunoassay, and enzyme-linked immunosorbent assay. However, none of these assays has proved to be sufficiently sensitive, specific, and reproducible to be used as a diagnostic test for NANBH.

Previously there was neither clarity nor agreement as to the identity or specificity of the antigen antibody systems associated with agents of NANBH. This was due, at least in part, to the prior or co-infection of HBV with NANBV in individuals, and to the known complexity of the soluble and particulate antigens associated with HBV, as well as to the integration of HBV DNA into the genome of liver cells. In addition, there is the possibility that NANBH is caused by more than one infectious agent, as well as the possibility that NANBH has been misdiagnosed. Moreover, it is unclear what the serological assays detect in the serum of patients with NANBH. It has been postulated that the agar-gel diffusion and counterimmunoelectrophoresis assays detect autoimmune responses or nonspecific protein interactions that sometimes occur between serum specimens, and that they do not represent specific NANBV antigen-antibody reactions. The immunofluorescence, and enzyme-linked immunosorbent, and radioimmunoassays appear to detect low levels of a rheumatoid-factor-like material that is frequently present in the serum of patients with NANBH as well as in patients with other hepatic and nonhepatic diseases. Some of the reactivity detected may represent antibody to host-determined cytoplasmic antigens.

There have been a number of candidate NANBV. See, for example the reviews by Prince (1983), Feinstone and Hoofnagle (1984), and Overby (1985, 1986, 1987) and the article by Iwarson (1987). However, there is no proof

that any of these candidates represent the etiological agent of NANBH.

The demand for sensitive, specific methods for screening and identifying carriers of NANBV and NANBV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and NANBH accounts for up to 90% of these cases. The major problem in this disease is the frequent progression to chronic liver damage (25-55%).

Patient care as well as the prevention of transmission of NANBH by blood and blood products or by close personal contact require reliable screening, diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to NANBV. In addition, there is also a need for effective vaccines and immunotherapeutic therapeutic agents for the prevention and/or treatment of the disease.

Applicant discovered a new virus, the Hepatitis C virus (HCV), which has proven to be the major etiologic agent of blood-borne NANBH (BB-NANBH). Applicant's initial work, including a partial genomic sequence of the prototype HCV isolate, CDC/HCV1 (also called HCV1), is described in EPO Pub. No. 318,216 (published 31 May 1989) and PCT Pub. No. WO 89/04669 (published 1 June 1989). The disclosures of these patent applications, as well as any corresponding national patent applications, are incorporated herein by reference. These applications teach, inter alia, recombinant DNA methods of cloning and expressing HCV sequences, HCV polypeptides, HCV immunodiagnostic techniques, HCV probe diagnostic techniques, anti-HCV antibodies, and methods of isolating new hCV sequences, including sequences of new HCV isolates.

Disclosure of the Invention

The present invention is based, in part, on new HCV sequences and polypeptides that are not disclosed in EPO Pub. No. 318,216, or in PCT Pub. No. WO 89/04669. Included within the invention is the application of these new sequences and polypeptides in, inter alia, immunodiagnostics, probe diagnostics, anti-HCV antibody production, PCR technology and recombinant DNA technology. Included within the invention, also, are new immunoassays based upon the immunogenicity of HCV polypeptides disclosed herein. The new subject matter claimed herein, while developed using techniques described in, for example, EPO Pub. No. 318,216, has a priority date which antecedes that publication, or any counterpart thereof. Thus, the invention provides novel compositions and methods useful for screening samples for HCV antigens and antibodies, and useful for treatment of HCV infections.

Accordingly, one aspect of the invention is a recombinant polynucleotide comprising a sequence derived from HCV cDNA, wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

Another aspect of the invention is a purified polypeptide comprising an epitope encoded within HCV cDNA wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

Yet another aspect of the invention is an immunogenic polypeptide produced by a cell transformed with a recombinant expression vector comprising an ORF of DNA derived from HCV cDNA, wherein the HCV cDNA is comprised

of a sequence derived from the HCV cDNA sequence in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the ORF is operably linked to a control sequence compatible with a desired host.

Another aspect of the invention is a peptide comprising an HCV epitope, wherein the peptide is of the formula

$$AA_x\text{-}AA_y,$$

wherein x and y designate amino acid numbers shown in Fig. 17, and wherein the peptide is selected from the group consisting of AA1-AA25, AA1-AA50, AA1-AA84, AA9-AA177, AA1-AA10, AA5-AA20, AA20-AA25, AA35-AA45, AA50-AA100, AA40-AA90, AA45-AA65, AA65-AA75, AA80-90, AA99-AA120, AA95-AA110, AA105-AA120, AA100-AA150, AA150-AA200, AA155-AA170, AA190-AA210, AA200-AA250, AA220-AA240, AA245-AA265, AA250-AA300, AA290-AA330, AA290-305, AA300-AA350, AA310-AA330, AA350-AA400, AA380-AA395, AA405-AA495, AA400-AA450, AA405-AA415, AA415-AA425, AA425-AA435, AA437-AA582, AA450-AA500, AA440-AA460, AA460-AA470, AA475-AA495, AA500-AA550, AA511-AA690, AA515-AA550, AA550-AA600, AA550-AA625, AA575-AA605, AA585-AA600, AA600-AA650, AA600-AA625, AA635-AA665, AA650-AA700, AA645-AA680, AA700-AA750, AA700-AA725, AA700-AA750, AA725-AA775, AA770-AA790, AA750-AA800, AA800-AA815, AA825-AA850, AA850-AA875, AA800-AA850, AA920-AA990, AA850-AA900, AA920-AA945, AA940-AA965, AA970-AA990, AA950-AA1000, AA1000-AA1060, AA1000-AA1025, AA1000-AA1050, AA1025-AA1040, AA1040-AA1055, AA1075-AA1175, AA1050-AA1200, AA1070-AA1100, AA1100-AA1130, AA1140-AA1165, AA1192-AA1457, AA1195-AA1250,

AA1200-AA1225, AA1225-AA1250, AA1250-AA1300, AA1260-AA1310, AA1260-AA1280, AA1266-AA1428, AA1300-AA1350, AA1290-AA1310, AA1310-AA1340, AA1345-AA1405, AA1345-AA1365, AA1350-AA1400, AA1365-AA1380, AA1380-AA1405, AA1400-AA1450, AA1450-AA1500, AA1460-AA1475, AA1475-AA1515, AA1475-AA1500, AA1500-AA1550, AA1500-AA1515, AA1515-AA1550, AA1550-AA1600, AA1545-AA1560, AA1569-AA1931, AA1570-AA1590, AA1595-AA1610, AA1590-AA1650, AA1610-AA1645, AA1650-AA1690, AA1685-AA1770, AA1689-AA1805, AA1690-AA1720, AA1694-AA1735, AA1720-AA1745, AA1745-AA1770, AA1750-AA1800, AA1775-AA1810, AA1795-AA1850, AA1850-AA1900, AA1900-AA1950, AA1900-AA1920, AA1916-AA2021, AA1920-AA1940, AA1949-AA2124, AA1950-AA2000, AA1950-AA1985, AA1980-AA2000, AA2000-AA2050, AA2005-AA2025, AA2020-AA2045, AA2045-AA2100, AA2045-AA2070, AA2054-AA2223, AA2070-AA2100, AA2100-AA2150, AA2150-AA2200, AA2200-AA2250, AA2200-AA2325, AA2250-AA2330, AA2255-AA2270, AA2265-AA2280, AA2280-AA2290, AA2287-AA2385, AA2300-AA2350, AA2290-AA2310, AA2310-AA2330, AA2330-AA2350, AA2350-AA2400, AA2348-AA2464, AA2345-AA2415, AA2345-AA2375, AA2370-AA2410, AA2371-AA2502, AA2400-AA2450, AA2400-AA2425, AA2415-AA2450, AA2445-AA2500, AA2445-AA2475, AA2470-AA2490, AA2500-AA2550, AA2505-AA2540, AA2535-AA2560, AA2550-AA2600, AA2560-AA2580, AA2600-AA2650, AA2605-AA2620, AA2620-AA2650, AA2640-AA2660, AA2650-AA2700, AA2655-AA2670, AA2670-AA2700, AA2700-AA2750, AA2740-AA2760, AA2750-AA2800, AA2755-AA2780, AA2780-AA2830, AA2785-AA2810, AA2796-AA2886, AA2810-AA2825, AA2800-AA2850, AA2850-AA2900,

AA2850-AA2865, AA2885-AA2905, AA2900-AA2950, AA2910-AA2930, AA2925-AA2950, AA2945-end(C' terminal).

Still another aspect of the invention is a monoclonal antibody directed against an epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Yet another aspect of the invention is a preparation of purified polyclonal antibodies directed against a polypeptide comprised of an epitope encoded within HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Still another aspect of the invention is a polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or from the complement of the HCV cDNA sequence.

Yet another aspect of the invention is a kit for analyzing samples for the presence of polynucleotides from HCV comprising a polynucleotide probe containing a nucleotide sequence of about 8 or more nucleotides, wherein the nucleotide sequence is derived from HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, wherein the polynucleotide probe is in a suitable container.

Another aspect of the invention is a kit for analyzing samples for the presence of an HCV antigen comprising an antibody which reacts immunologically with an HCV antigen, wherein the antigen contains an epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Yet another aspect of the invention is a kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide containing an HCV epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Another aspect of the invention is a kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide expressed from HCV cDNA in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, wherein the antigenic polypeptide is present in a suitable container.

Still another aspect of the invention is a method for detecting HCV nucleic acids in a sample comprising:

(a) reacting nucleic acids of the sample with a polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, and wherein the

reacting is under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample; and (b) detecting a polynucleotide duplex which contains the probe, formed in step (a).

Yet another aspect of the invention is an immunoassay for detecting an HCV antigen comprising:

(a) incubating a sample suspected of containing an HCV antigen with an antibody directed against an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and (b) detecting an antibody-antigen complex formed in step (a) which contains the antibody.

Still another aspect of the invention is an immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing anti-HCV antibodies with an antigen polypeptide containing an epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and detecting an antibody-antigen complex formed in step (a) which contains the antigen polypeptide.

Another aspect of the invention is a vaccine for treatment of HCV infection comprising an immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17 or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

Yet another aspect of the invention is a method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polyeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is of the sequence present in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

Still another aspect of the invention is an antisense polynucleotide derived from HCV cDNA, wherein the HCV cDNA is that shown in Fig. 17.

Yet another aspect of the invention is a method for preparing purified fusion polypeptide C100-3 comprising:

(a) providing a crude cell lysate containing polypeptide C100-3,

(b) treating the crude cell lysate with an amount of acetone which causes the polypeptide to precipitate,

(c) isolating and solubilizing the precipitated material,

(d) isolating the C100-3 polypeptide by anion exchange chromatography, and

(e) further isolating the C100-3 polypeptide of step (d) by gel filtration.

Brief Description of the Drawings

Fig. 1 shows the sequence of the HCV cDNA in clone 12f, and the amino acids encoded therein.

Fig. 2 shows the HCV cDNA sequence in clone k9-1, and the amino acids encoded therein.

Fig. 3 shows the sequence of clone 15e, and the amino acids encoded therein.

Fig. 4 shows the nucleotide sequence of HCV cDNA in clone 13i, the amino acids encoded therein, and the sequences which overlap with clone 12f.

Fig. 5 shows the nucleotide sequence of HCV cDNA in clone 26j, the amino acids encoded therein, and the sequences which overlap clone 13i.

Fig. 6 shows the nucleotide sequence of HCV cDNA in clone CA59a, the amino acids encoded therein, and the sequences which overlap with clones 26j and K9-1.

Fig. 7 shows the nucleotide sequence of HCV cDNA in clone CA84a, the amino acids encoded therein, and the sequences which overlap with clone CA59a.

Fig. 8 shows the nucleotide sequence of HCV cDNA in clone CA156e, the amino acids encoded therein, and the sequences which overlap with CA84a.

Fig. 9 shows the nucleotide sequence of HCV cDNA in clone CA167b, the amino acids encoded therein, and the sequences which overlap CA156e.

Fig. 10 shows the nucleotide sequence of HCV cDNA in clone CA216a, the amino acids encoded therein, and the overlap with clone CA167b.

Fig. 11 shows the nucleotide sequence of HCV cDNA in clone CA290a, the amino acids encoded therein, and the overlap with clone CA216a.

Fig. 12 shows the nucleotide sequence of HCV cDNA in clone ag30a and the overlap with clone CA290a.

Fig. 13 shows the nucleotide sequence of HCV cDNA in clone CA205a, and the overlap with the HCV cDNA sequence in clone CA290a.

Fig. 14 shows the nucleotide sequence of HCV cDNA in clone 18g, and the overlap with the HCV cDNA sequence in clone ag30a.

Fig. 15 shows the nucleotide sequence of HCV cDNA in clone 16jh, the amino acids encoded therein, and the overlap of nucleotides with the HCV cDNA sequence in clone 15e.

Fig. 16 shows the ORF of HCV cDNA derived from clones pi14a, CA167b, CA156e, CA84a, CA59a, K9-1, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, and 15e.

Fig. 17 shows the sense strand of the compiled HCV cDNA sequence derived from the above-described clones and the compiled HCV cDNA sequence published in EPO Pub. No. 318,216. The clones from which the sequence was derived are b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, CA59a, K9-1 (also called k9-1),26j, 13i, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a, and 16jh. In the figure the three horizontal dashes above the sequence indicate the position

of the putative initiator methionine codon; the two vertical dashes indicate the first and last nucleotides of the published sequence. Also shown in the figure is the amino acid sequence of the putative polyprotein encoded in the HCV cDNA.

Fig. 18 is a diagram of the immunological colony screening method used in antigenic mapping studies.

Fig. 19 shows the hydrophobicity profiles of polyproteins encoded in HCV and in West Nile virus.

Fig. 20 is a tracing of the hydrophilicity/ hydrophobicity profile and of the antigenic index of the putative HCV polyprotein.

Fig. 21 shows the conserved co-linear peptides in HCV and Flaviviruses.

Modes for Carrying Out the Invention

I. Definitions

The term "hepatitis C virus" has been reserved by workers in the field for an heretofore unknown etiologic agent of NANBH. Accordingly, as used herein, "hepatitis C virus" (HCV) refers to an agent causitive of NANBH, which was formerly referred to as NANBV and/or BB-NANBV. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein. As an extension of this terminology, the disease caused by HCV, formerly called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably herein.

The term "HCV", as used herein, denotes a viral species of which pathogenic strains cause NANBH, and attenuated strains or defective interfering particles derived therefrom. As shown infra., the HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly on the order of 10^{-3} to 10^{-4} per incorporated

nucleotide (Fields & Knipe (1986)). Therefore, there are multiple strains, which may be virulent or avirulent, within the HCV species described infra. The compositions and methods described herein, enable the propagation, identification, detection, and isolation of the various HCV strains or isolates. Moreover, the disclosure herein allows the preparation of diagnostics and vaccines for the various strains, as well as compositions and methods that have utility in screening procedures for anti-viral agents for pharmacologic use, such as agents that inhibit replication of HCV.

The information provided herein, although derived from the prototype strain or isolate of HCV, hereinafter referred to as CDC/HCV1 (also called HCV1), is sufficient to allow a viral taxonomist to identify other strains which fall within the species. The information provided herein allows the belief that HCV is a Flavi-like virus. The morphology and composition of Flavivirus particles are known, and are discussed in Brinton (1986). Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

Different strains or isolates of HCV are expected to contain variations at the amino acid and nucleic acids compared with the prototype isolate, HCV1. Many isolates are expected to show much (i.e. more than about 40%) homology in the total amino acid sequence compared with HCV1. However, it may also be found that other less homologous HCV isolates. These would be defined as HCV strains according to various criteria such as an ORF of approximately 9,000 nucleotides to ap-

proximately 12,000 nucleotides, encoding a polyprotein similar in size to that of HCV1, an encoded polyprotein of similar hydrophobic and antigenic character to that of HCV1, and the presence of co-linear peptide sequences that are conserved with HCV1. In addition, the genome would be a positive-stranded RNA.

HCV encodes at least one epitope which is immunologically identifiable with an epitope in the HCV genome from which the cDNAs described herein are derived; preferably the epitope is contained an amino acid sequence described herein. The epitope is unique to HCV when compared to other known Flaviviruses. The uniqueness of the epitope may be determined by its immunological reactivity with anti-HCV antibodies and lack of immunological reactivity with antibodies to other Flavivirus species. Methods for determining immunological reactivity are known in the art, for example, by radioimmunoassay, by Elisa assay, by hemagglutination, and several examples of suitable techniques for assays are provided herein.

In addition to the above, the following parameters of nucleic acid homology and amino acid homology are applicable, either alone or in combination, in identifying a strain or isolate as HCV. Since HCV strains and isolates are evolutionarily related, it is expected that the overall homology of the genomes at the nucleotide level probably will be about 40% or greater, probably about 60% or greater, and even more probably about 80% or greater; and in addition that there will be corresponding contiguous sequences of at least about 13 nucleotides. The correspondence between the putative HCV strain genomic sequence and the CDC/HCV1 cDNA sequence can be determined by techniques known in the art. For example, they can be determined by a direct comparison of the sequence information of the polynucleotide from the putative HCV, and the HCV cDNA sequence(s) described herein. For example, also,

they can be determined by hybridization of the polynucleotides under conditions which form stable duplexes between homologous regions (for example, those which would be used prior to S_1 digestion), followed by digestion with single stranded specific nuclease(s), followed by size determination of the digested fragments.

Because of the evolutionary relationship of the strains or isolates of HCV, putative HCV strains or isolates are identifiable by their homology at the polypeptide level. Generally, HCV strains or isolates are expected to be more than about 40% homologous, probably more than about 70% homologous, and even more probably more than about 80% homologous, and some may even be more than about 90% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to the sequences provided herein. Alternatively the nucleotide sequence of the genomic material of the putative HCV may be determined (usually via a cDNA intermediate), the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

As used herein, a polynucleotide "derived from" a designated sequence refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6 nucleotides, preferably at least about 8 nucleotides, more preferably at least about 10-12 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding to a region of the designated nucleotide sequence. "Corresponding" means homologous to or complementary to the designated sequence. Preferably, the sequence of the region from which the polynucleotide is derived is homologous to or complementary to a sequence which is unique to an HCV genome. Whether or not a sequence is unique to the HCV genome can be determined by

techniques known to those of skill in the art. For example, the sequence can be compared to sequences in databanks, e.g., Genebank, to determine whether it is present in the uninfected host or other organisms. The sequence can also be compared to the known sequences of other viral agents, including those which are known to induce hepatitis, e.g., HAV, HBV, and HDV, and to other members of the Flaviviridae. The correspondence or non-correspondence of the derived sequence to other sequences can also be determined by hybridization under the appropriate stringency conditions. Hybridization techniques for determining the complementarity of nucleic acid sequences are known in the art, and are discussed infra. See also, for example, Maniatis et al. (1982). In addition, mismatches of duplex polynucleotides formed by hybridization can be determined by known techniques, including for example, digestion with a nuclease such as S1 that specifically digests single-stranded areas in duplex polynucleotides. Regions from which typical DNA sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

The derived polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be consistent with an intended use.

Similarly, a polypeptide or amino acid sequence "derived from" a designated nucleic acid sequence refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least

3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence.

A recombinant or derived polypeptide is not necessarily translated from a designated nucleic acid sequence, for example, the HCV cDNA sequences described herein, or from an HCV genome; it may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from mutated HCV. A recombinant or derived polypeptide may include one or more analogs of amino acids or unnatural amino acids in its sequence. Methods of inserting analogs of amino acids into a sequence are known in the art. It also may include one or more labels, which are known to those of skill in the art.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation which: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, or (3) does not occur in nature.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, as well as double- and single stranded RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl

phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

The term "purified viral polynucleotide" refers to an HCV genome or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90% of polypeptides with which the viral polynucleotide is naturally associated. Techniques for purifying viral polynucleotides from viral particles are known in the art, and include for example, disruption of the particle with a chaotropic agent, differential extraction and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

The term "purified viral polypeptide" refers to an HCV polypeptide or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90%, of cellular components with which the viral polypeptide is naturally associated. Techniques for purifying viral polypeptides are known in the art, and examples of these techniques are discussed infra. The term "purified viral polynucleotide" refers to an HCV genome or fragment thereof which is essentially free, i.e., contains less than about 20%, preferably less than about 50%, and

even more preferably less than about 70% of polypeptides with which the viral polynucleotide is naturally associated. Techniques for purifying viral polynucleotides from viral particles are known in the art, and include for example, disruption of the particle with a chaotropic agent, and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc. that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers.

The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

"Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are also present in the designated polypeptide(s), usually HCV proteins. Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art, and are also illustrated infra.

As used herein, "epitope" refers to an antigenic determinant of a polypeptide; an epitope could comprise 3 amino acids in a spatial conformation which is unique to

the epitope, generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance.

A polypeptide is "immunologically reactive" with an antibody when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an epitope against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art.

As used herein, the term "immunogenic polypeptide" is a polypeptide that elicits a cellular and/or humoral response, whether alone or linked to a carrier in the presence or absence of an adjuvant.

The term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

"Treatment" as used herein refers to prophylaxis and/or therapy.

An "individual", as used herein, refers to vertebrates, particularly members of the mammalian species, and includes but is not limited to domestic animals, sports animals, and primates, including humans.

As used herein, the "sense strand" of a nucleic acid contains the sequence that has sequence homology to that of mRNA. The "anti-sense strand" contains a sequence which is complementary to that of the "sense strand".

As used herein, a "positive stranded genome" of a virus is one in which the genome, whether RNA or DNA, is single-stranded and which encodes a viral polypeptide(s). Examples of positive stranded RNA viruses include Togaviridae, Coronaviridae, Retroviridae, Picornaviridae, and Caliciviridae. Included also, are the Flaviviridae, which were formerly classified as Togaviradae. See Fields & Knipe (1986).

As used herein, "antibody-containing body component" refers to a component of an individual's body which is a source of the antibodies of interest. Antibody containing body components are known in the art, and include but are not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, white blood cells, and myelomas.

As used herein, "purified HCV" refers to a preparation of HCV which has been isolated from the cellular constituents with which the virus is normally associated, and from other types of viruses which may be present in the infected tissue. The techniques for isolating viruses are known to those of skill in the art, and include, for example, centrifugation and affinity chromatography; a method of preparing purified HCV is discussed infra.

The term "HCV particles" as used herein include entire virion as well as particles which are intermediates in virion formation. HCV particles generally have one or more HCV proteins associated with the HCV nucleic acid.

As used herein, the term "probe" refers to a polynucleotide which forms a hybrid structure with a sequence in a target region, due to complementarity of at least one sequence in the probe with a sequence in the target region. The probe, however, does not contain a sequence complementary to sequence(s) used to prime the polymerase chain reaction.

As used herein, the term "target region" refers to a region of the nucleic acid which is to be amplified and/or detected.

As used herein, the term "viral RNA", which includes HCV RNA, refers to RNA from the viral genome, fragments thereof, transcripts thereof, and mutant sequences derived therefrom.

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of <u>in vitro</u> cell culture constituents (including but not limited to conditioned medium resulting from the

growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

II. Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, MOLECULAR CLONING; A LABORATORY MANUAL (1982); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

The useful materials and processes of the present invention are made possible by the provision of a family of nucleotide sequences isolated from cDNA libraries which contain HCV cDNA sequences. These cDNA librar-

ies were derived from nucleic acid sequences present in the plasma of an HCV-infected chimpanzee. The construction of one of these libraries, the "c" library (ATCC No. 40394), was reported in EPO Pub. No. 318,216. Several of the clones containing HCV cDNA reported herein were obtained from the "c" library. Although other clones reported herein were obtained from other HCV cDNA libraries, the presence of clones containing the sequences in the "c" library was confirmed. As discussed in EPO Pub. No. 318,216, the family of HCV cDNA sequences isolated from the "c" library are not of human or chimpanzee origin, and show no significant homology to sequences contained within the HBV genome.

The availability of the HCV cDNAs described herein permits the construction of polynucleotide probes which are reagents useful for detecting viral polynucleotides in biological samples, including donated blood. For example, from the sequences it is possible to synthesize DNA oligomers of about 8-10 nucleotides, or larger, which are useful as hybridization probes to detect the presence of HCV RNA in, for example, donated blood, sera of subjects suspected of harboring the virus, or cell culture systems in which the virus is replicating. In addition, the cDNA sequences also allow the design and production of HCV specific polypeptides which are useful as diagnostic reagents for the presence of antibodies raised during HCV infection. Antibodies to purified polypeptides derived from the cDNAs may also be used to detect viral antigens in biological samples, including, for example, donated blood samples, sera from patients with NANBH, and in tissue culture systems being used for HCV replication. Moreover, the immunogenic polypeptides disclosed herein, which are encoded in portions of the ORF of HCV cDNA shown in Fig. 17, are also useful for HCV

screening, diagnosis, and treatment, and for raising antibodies which are also useful for these purposes.

In addition, the novel cDNA sequences described herein enable further characterization of the HCV genome. Polynucleotide probes and primers derived from these sequences may be used to amplify sequences present in cDNA libraries, and/or to screen cDNA libraries for additional overlapping cDNA sequences, which, in turn, may be used to obtain more overlapping sequences. As indicated infra. and in EPO Pub. No. 318,216, the genome of HCV appears to be RNA comprised primarily of a large open reading frame (ORF) which encodes a large polyprotein.

The HCV cDNA sequences provided herein, the polypeptides derived from these sequences, and the immunogenic polypeptides described herein, as well as antibodies directed against these polypeptides are also useful in the isolation and identification of the blood-borne NABV (BB-NANBV) agent(s). For example, antibodies directed against HCV epitopes contained in polypeptides derived from the cDNAs may be used in processes based upon affinity chromatography to isolate the virus. Alternatively, the antibodies may be used to identify viral particles isolated by other techniques. The viral antigens and the genomic material within the isolated viral particles may then be further characterized.

In addition to the above, the information provided infra allows the identification of additional HCV strains or isolates. The isolation and characterization of the additional HCV strains or isolates may be accomplished by isolating the nucleic acids from body components which contain viral particles and/or viral RNA, creating cDNA libraries using polynucleotide probes based on the HCV cDNA probes described infra., screening the libraries for clones containing HCV cDNA sequences described infra., and comparing the HCV cDNAs from the new

isolates with the cDNAs described infra. The polypeptides encoded therein, or in the viral genome, may be monitored for immunological cross-reactivity utilizing the polypeptides and antibodies described supra. Strains or isolates which fit within the parameters of HCV, as described in the Definitions section, supra., are readily identifiable. Other methods for identifying HCV strains will be obvious to those of skill in the art, based upon the information provided herein.

Isolation of the HCV cDNA Sequences

The novel HCV cDNA sequences described infra. extend the sequence of the cDNA to the HCV genome reported in EPO Pub. No. 318,216. The sequences which are present in clones b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, and CA59a lie upstream of the reported sequence, and when compiled, yield nucleotides nos. -319 to 1348 of the composite HCV cDNA sequence. (The negative number on a nucleotide indicates its distance upstream of the nucleotide which starts the putative initiator MET codon.) The sequences which are present in clones b5a and 16jh lie downstream of the reported sequence, and yield nucleotides nos. 8659 to 8866 of the composite sequence. The composite HCV cDNA sequence which includes the sequences in the aforementioned clones, is shown in Fig. 17.

The novel HCV cDNAs described herein were isolated from a number of HCV cDNA libraries, including the "c" library present in lambda gt11 (ATCC No. 40394). The HCV cDNA libraries were constructed using pooled serum from a chimpanzee with chronic HCV infection and containing a high titer of the virus, i.e., at least 10^6 chimp infectious doses/ml (CID/ml). The pooled serum was used to isolate viral particles; nucleic acids isolated from these particles was used as the template in the construc-

tion of cDNA libraries to the viral genome. The procedures for isolation of putative HCV particles and for constructing the "c" HCV cDNA library is described in EPO Pub. No. 318,216. Other methods for constructing HCV cDNA libraries are known in the art, and some of these methods are described infra., in the Examples. Isolation of the sequences was by screening the libraries using synthetic polynucleotide probes, the sequences of which were derived from the 5'-region and the 3'-region of the known HCV cDNA sequence. The description of the method to retrive the cDNa sequences is mostly of historical interest. The resultant sequences (and their complements) are provided herein, and the sequences, or any portion thereof, could be prepared using synthetic methods, or by a combination of synthetic methods with retrieval of partial sequences using methods similar to those described herein.

Preparation of Viral Polypeptides and Fragments

The availability of HCV cDNA sequences, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding antigenically active regions of the polypeptide encoded in either strand. These antigenically active regions may be derived from coat or envelope antigens or from core antigens, or from antigens which are non-structural including, for example, polynucleotide binding proteins, polynucleotide polymerase(s), and other viral proteins required for the replication and/or assembly of the virus particle. Fragments encoding the desired polypeptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors

which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986. Vectors for the expression of fusion polypeptides of SOD and HCV polypeptides encoded in a number of HCV clones are described infra., in the Examples. Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant polypeptide, such as a mature or fusion protein; alternatively, a polypeptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell lines is given infra. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such polypeptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, as discussed infra., antibodies to these polypeptides are useful for isolating and identifying HCV particles.

Preparation of Antigenic Polypeptides and Conjugation with Carrier

An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of HCV antigen. Accordingly, using the cDNAs of HCV as a basis, DNAs encoding short segments of HCV polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the synthesized polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun. Rev. (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimidomethyl)cyclohexane-1-carboxylic acid, and the like. The

carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

In addition to full-length viral proteins, polypeptides comprising truncated HCV amino acid sequences encoding at least one viral epitope are useful immunological reagents. For example, polypeptides comprising such truncated sequences can be used as reagents in an immunoassay. These polypeptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While these truncated sequences can be produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant polypeptides comprising an HCV sequence. Polypeptides comprising these truncated HCV sequences can be made up entirely of HCV sequences (one or more epitopes, either contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein.

Useful heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, e.g., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783, the disclosures of which are incorporated herein by reference.

The size of polypeptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

Truncated HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire viral protein sequence can be screened by preparing a series of short peptides that together span the entire protein sequence. An example of antigenic screening of the regions of the HCV polyprotein is shown infra. In addition, by starting with, for example, 100mer polypeptides, it would be routine to test each polypeptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified 100mer to map the epitope of interest. Screening such peptides in an

immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare oligopeptides comprising the identified regions for screening. Such a computer analysis of the HCV amino acid sequence is shown in Fig. 20, where the hydrophilic/ hydrophobic character is displayed above the antigen index. The amino acids are numbered from the starting MET (position 1) as shown in Fig. 17. It is appreciated by those of skill in the art that such computer analysis of antigenicity does not always identify an epitope that actually exists, and can also incorrectly identify a region of the protein as containing an epitope.

Examples of HCV amino acid sequences that may be useful, which are expressed from expression vectors comprised of clones 5-1-1, 81, CA74a, 35f, 279a, C36, C33b, CA290a, C8f, C12f, 14c, 15e, C25c, C33c, C33f, 33g, C39c, C40b, CA167b are described infra. Other examples of HCV amino acid sequences that may be useful as described herein are set forth below. It is to be understood that these peptides do not necessarily precisely map one epitope, and may also contain HCV sequence that is not immunogenic. These non-immunogenic portions of the sequence can be defined as described above using conventional techniques and deleted from the described sequences. Further, additional truncated HCV amino acid sequences that comprise an epitope or are immunogenic can be identified as described above. The following sequences are given by amino acid number (i.e., "AAn") where n is the amino acid number as shown in Fig. 17:

AA1-AA25; AA1-AA50; AA1-AA84; AA9-AA177; AA1-AA10; AA5-AA20; AA20-AA25; AA35-AA45; AA50-AA100; AA40-AA90; AA45-AA65; AA65-AA75; AA80-90; AA99-AA120; AA95-AA110; AA105-AA120; AA100-AA150; AA150-AA200; AA155-AA170; AA190-AA210; AA200-AA250; AA220-AA240;

AA245-AA265; AA250-AA300; AA290-AA330; AA290-305;
AA300-AA350; AA310-AA330; AA350-AA400; AA380-AA395;
AA405-AA495; AA400-AA450; AA405-AA415; AA415-AA425;
AA425-AA435; AA437-AA582; AA450-AA500; AA440-AA460;
AA460-AA470; AA475-AA495; AA500-AA550; AA511-AA690;
AA515-AA550; AA550-AA600; AA550-AA625; AA575-AA605;
AA585-AA600; AA600-AA650; AA600-AA625; AA635-AA665;
AA650-AA700; AA645-AA680; AA700-AA750; AA700-AA725;
AA700-AA750; AA725-AA775; AA770-AA790; AA750-AA800;
AA800-AA815; AA825-AA850; AA850-AA875; AA800-AA850;
AA920-AA990; AA850-AA900; AA920-AA945; AA940-AA965;
AA970-AA990; AA950-AA1000; AA1000-AA1060;
AA1000-AA1025; AA1000-AA1050; AA1025-AA1040;
AA1040-AA1055; AA1075-AA1175; AA1050-AA1200;
AA1070-AA1100; AA1100-AA1130; AA1140-AA1165;
AA1192-AA1457; AA1195-AA1250; AA1200-AA1225;
AA1225-AA1250; AA1250-AA1300; AA1260-AA1310;
AA1260-AA1280; AA1266-AA1428; AA1300-AA1350;
AA1290-AA1310; AA1310-AA1340; AA1345-AA1405;
AA1345-AA1365; AA1350-AA1400; AA1365-AA1380;
AA1380-AA1405; AA1400-AA1450; AA1450-AA1500;
AA1460-AA1475; AA1475-AA1515; AA1475-AA1500;
AA1500-AA1550; AA1500-AA1515; AA1515-AA1550;
AA1550-AA1600; AA1545-AA1560; AA1569-AA1931;
AA1570-AA1590; AA1595-AA1610; AA1590-AA1650;
AA1610-AA1645; AA1650-AA1690; AA1685-AA1770;
AA1689-AA1805; AA1690-AA1720; AA1694-AA1735;
AA1720-AA1745; AA1745-AA1770; AA1750-AA1800;
AA1775-AA1810; AA1795-AA1850; AA1850-AA1900;
AA1900-AA1950; AA1900-AA1920; AA1916-AA2021;
AA1920-AA1940; AA1949-AA2124; AA1950-AA2000;
AA1950-AA1985; AA1980-AA2000; AA2000-AA2050;
AA2005-AA2025; AA2020-AA2045; AA2045-AA2100;
AA2045-AA2070; AA2054-AA2223; AA2070-AA2100;
AA2100-AA2150; AA2150-AA2200; AA2200-AA2250;
AA2200-AA2325; AA2250-AA2330; AA2255-AA2270;
AA2265-AA2280; AA2280-AA2290; AA2287-AA2385;
AA2300-AA2350; AA2290-AA2310; AA2310-AA2330;
AA2330-AA2350; AA2350-AA2400; AA2348-AA2464;
AA2345-AA2415; AA2345-AA2375; AA2370-AA2410;
AA2371-AA2502; AA2400-AA2450; AA2400-AA2425;
AA2415-AA2450; AA2445-AA2500; AA2445-AA2475;
AA2470-AA2490; AA2500-AA2550; AA2505-AA2540;
AA2535-AA2560; AA2550-AA2600; AA2560-AA2580;
AA2600-AA2650; AA2605-AA2620; AA2620-AA2650;
AA2640-AA2660; AA2650-AA2700; AA2655-AA2670;
AA2670-AA2700; AA2700-AA2750; AA2740-AA2760;
AA2750-AA2800; AA2755-AA2780;
AA2780-AA2830; AA2785-AA2810; AA2796-AA2886;
AA2810-AA2825; AA2800-AA2850; AA2850-AA2900;
AA2850-AA2865; AA2885-AA2905; AA2900-AA2950;

AA2910-AA2930; AA2925-AA2950; AA2945-end(C' terminal).

The above HCV amino acid sequences can be prepared as discrete peptides or incorporated into a larger polypeptide, and may find use as described herein. Additional polypeptides comprising truncated HCV sequences are described in the examples.

The observed relationship of the putative polyproteins of HCV and the Flaviviruses allows some prediction of the putative domains of the HCV "non-structural" (NS) proteins. The locations of the individual NS proteins in the putative Flavivirus precursor polyprotein are fairly well-known. Moreover, these also coincide with observed gross fluctuations in the hydrophobicity profile of the polyprotein. It is established that NS5 of Flaviviruses encodes the virion polymerase, and that NS1 corresponds with a complement fixation antigen which has been shown to be an effective vaccine in animals. Recently, it has been shown that a flaviviral protease function resides in NS3. Due to the observed similarities betwen HCV and the Flaviviruses, described infra., deductions concerning the approximate locations of the corresponding protein domains and functions in the HCV polyprotein are possible. The expression of polypeptides containing these domains in a variety of recombinant host cells, including, for example, bacteria, yeast, insect, and vertebrate cells, should give rise to important immunological reagents which can be used for diagnosis, detection, and vaccines.

Although the non-structural protein regions of the putative polyproteins of the HCV isolate described herein and of Flaviviruses appear to have some similarity, there is less similarity between the putative structural regions which are towards the N-terminus. In this region, there is a greater divergence in sequence, and in addi-

tion, the hydrophobic profile of the two regions show less similarity. This "divergence" begins in the N-terminal region of the putative NS1 domain in HCV, and extends to the presumed N-terminus. Nevertheless, it may still be possible to predict the approximate locations of the putative nucleocapsid (N-terminal basic domain) and E (generally hydrophobic) domains within the HCV polyprotein. In the Examples the predictions are based on the changes observed in the hydrophobic profile of the HCV polyprotein, and on a knowledge of the location and character of the flaviviral proteins. From these predictions it may be possible to identify approximate regions of the HCV polyprotein that could correspond with useful immunological reagents. For example, the E and NS1 proteins of Flaviviruses are known to have efficacy as protective vaccines. These regions, as well as some which are shown to be antigenic in the HCV isolate described herein, for example those within putative NS3, C, and NS5, etc., should also provide diagnostic reagents. Moreover, the location and expression of viral-encoded enzymes may also allow the evaluation of anti-viral enzyme inhibitors, i.e., for example, inhibitors which prevent enzyme activity by virtue of an interaction with the enzyme itself, or substances which may prevent expression of the enzyme, (for example, anti-sense RNA, or other drugs which interfere with expression).

Preparation of Hybrid Particle Immunogens Containing HCV Epitopes

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. Constructs wherein the NANBV epitope is linked directly to the particle-forming protein

coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S. cerevisiae (Valenzuela et al. (1982)), as well as in, for example, mammalian cells (Valenzuela, P., et al. (1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

Preparation of Vaccines

Vaccines may be prepared from one or more immunogenic polypeptides derived from HCV cDNA, including the cDNA sequences described in the Examples. The

observed homology between HCV and Flaviviruses provides information concerning the polypeptides which may be most effective as vaccines, as well as the regions of the genome in which they are encoded. The general structure of the Flavivirus genome is discussed in Rice et al (1986). The flavivirus genomic RNA is believed to be the only virus-specific mRNA species, and it is translated into the three viral structural proteins, i.e., C, M, and E, as well as two large nonstructural proteins, NS4 and NS5, and a complex set of smaller nonstructural proteins. It is known that major neutralizing epitopes for Flaviviruses reside in the E (envelope) protein (Roehrig (1986)). Thus, vaccines may be comprised of recombinant polypeptides containing epitopes of HCV E. These polypeptides may be expressed in bacteria, yeast, or mammalian cells, or alternatively may be isolated from viral preparations. It is also anticipated that the other structural proteins may also contain epitopes which give rise to protective anti-HCV antibodies. Thus, polypeptides containing the epitopes of E, C, and M may also be used, whether singly or in combination, in HCV vaccines.

In addition to the above, it has been shown that immunization with NS1 (nonstructural protein 1), results in protection against yellow fever (Schlesinger et al (1986)). This is true even though the immunization does not give rise to neutralizing antibodies. Thus, particularly since this protein appears to be highly conserved among Flaviviruses, it is likely that HCV NS1 will also be protective against HCV infection. Moreover, it also shows that nonstructural proteins may provide protection against viral pathogenicity, even if they do not cause the production of neutralizing antibodies.

The information provided in the Examples concerning the immunogenicity of the polypeptides

expressed from cloned HCV cDNAs which span the various regions of the HCV ORF also allows predictions concerning their use in vaccines.

In view of the above, multivalent vaccines against HCV may be comprised of one or more epitopes from one or more structural proteins, and/or one or more epitopes from one or more nonstructural proteins. These vaccines may be comprised of, for example, recombinant HCV polypeptides and/or polypeptides isolated from the virions. In particular, vaccines are contemplated comprising one or more of the following HCV proteins, or subunit antigens derived therefrom: E, NS1, C, NS2, NS3, NS4 and NS5. Particularly preferred are vaccines comprising E and/or NS1, or subunits thereof.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-

isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with

the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

Dosage and Administration of Vaccines

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic HCV antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

Preparation of Antibodies Against HCV Epitopes

The immunogenic polypeptides prepared as described above are used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Alternatively, polyclonal antibodies may be isolated from a mammal which has been previously infected with HCV. An example of a method for purifying antibodies to HCV epitopes from serum from an infected individual, based upon affinity chromatography and utilizing a fusion polypeptide of SOD and a polypeptide encoded within cDNA clone 5-1-1, is presented in EPO Pub. No. 318,216.

Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985).

Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for treatment and/or diagnosis of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

It would also be recognized by one of ordinary skill in the art that a variety of types of antibodies directed against HCV epitopes may be produced. As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one antibody combining site. An "antibody combining site" or "binding domain" is formed from the folding of variable domains of an antibody molecule(s) to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows an immunological reaction with the antigen. An antibody combining site may be formed from a heavy and/or a light chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, altered antibodies, univalent

antibodies, the Fab proteins, and single domain antibodies.

A "single domain antibody" (dAb) is an antibody which is comprised of an VH domain, which reacts immunologically with a designated antigen. A dAB does not contain a VL domain, but may contain other antigen binding domains known to exist in antibodies, for example, the kappa and lambda domains. Methods for preparing dABs are known in the art. See, for example, Ward et al. (1989).

Antibodies may also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation are known in the art (see, e.g., U.S. Patent No. 4,816,467, which is incorporated herein by reference), and include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of all the chains of a particular antibody are homologous with the chains found in one antibody produced by the lymphocyte which produces that antibody in situ, or in vitro (for example, in hybridomas). Vertebrate antibodies typicallly include native antibodies, for example, purified polyclonal antibodies and monoclonal antibodies. Examples of the methods for the preparation of these antibodies are described infra.

"Hybrid antibodies" are antibodies wherein one pair of heavy and light chains is homologous to those in a first antibody, while the other pair of heavy and light chains is homologous to those in a different second antibody. Typically, each of these two pairs will bind different epitopes, particularly on different antigens. This results in the property of "divalence", i.e., the ability

to bind two antigens simultaneously. Such hybrids may also be formed using chimeric chains, as set forth below.

"Chimeric antibodies", are antibodies in which the heavy and/or light chains are fusion proteins. Typically the constant domain of the chains is from one particular species and/or class, and the variable domains are from a different species and/or class. Also included is any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be differing classes, or different species of origin, and whether or not the fusion point is at the variable/constant boundary. Thus, it is possible to produce antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

Another example is "altered antibodies", which refers to antibodies in which the naturally occurring amino acid sequence in a vertebrate antibody has been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region may be made to alter antigen binding characeristics. The antibody may also be engineered to aid the specific delivery of a

molecule or substance to a specific cell or tissue site. The desired alterations may be made by known techniques in molecular biology, e.g., recombinant techniques, site directed mutagenesis, etc.

Yet another example are "univalent antibodies", which are aggregates comprised of a heavy chain/light chain dimer bound to the Fc (i.e., constant) region of a second heavy chain. This type of antibody escapes antigenic modulation. See, e.g., Glennie et al. (1982).

Included also within the definition of antibodies are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion . "Fab" includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as $F(ab)_2$), which are capable of selectively reacting with a designated antigen or antigen family. "Fab" antibodies may be divided into subsets analogous to those described above, i.e, "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing "Fab" fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

II.H. Diagnostic Oligonucleotide Probes and Kits

Using the disclosed portions of the isolated HCV cDNAs as a basis, oligomers of approximately 8 nucleotides or more can be prepared, either by excision or synthetically, which hybridize with the HCV genome and are useful in identification of the viral agent(s), further characterization of the viral genome(s), as well as in

detection of the virus(es) in diseased individuals. The probes for HCV polynucleotides (natural or derived) are a length which allows the detection of unique viral sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 20 nucleotides appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. Among useful probes, for example, are those derived from the newly isolated clones disclosed herein, as well as the various oligomers useful in probing cDNA libraries, set forth below. A complement to any unique portion of the HCV genome will be satisfactory. For use as probes, complete complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

For use of such probes as diagnostics, the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. The probes are then labeled. Suitable labels, and methods for labeling probes are known in the art, and include, for example, radioactive labels incorporated by nick translation or kinasing, biotin, fluorescent probes, and chemiluminescent probes. The nucleic acids extracted from the sample are then treated with the labeled probe under hybridization conditions of suitable stringencies, and polynucleotide duplexes containing the probe are detected.

The probes can be made completely complementary to the HCV genome. Therefore, usually high stringency conditions are desirable in order to prevent false

positives. However, conditions of high stringency should only be used if the probes are complementary to regions of the viral genome which lack heterogeneity. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and concentration of formamide. These factors are outlined in, for example, Maniatis, T. (1982).

Generally, it is expected that the HCV genome sequences will be present in serum of infected individuals at relatively low levels, i.e., at approximately 10^2-10^3 chimp infectious doses (CID) per ml. This level may require that amplification techniques be used in hybridization assays. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT-tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT application 84/03520 and EPA124221 describe a DNA hybridization assay in which: (1) analyte is annealed to a single-stranded DNA probe that is complementary to an enzyme-labeled oligonucleotide; and (2) the resulting tailed duplex is hybridized to an enzyme-labeled oligonucleotide. EPA 204510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands. A particularly desirable technique may first involve amplification of the target HCV sequences in sera approximately 10,000 fold, i.e., to approximately 10^6 sequences/ml. This may be accomplished, for example, by the polymerase chain reactions (PCR) technique described which is by Saiki et al. (1986),

by Mullis, U.S. Patent No. 4,683,195, and by Mullis et al. U.S. Patent No. 4,683,202. The amplified sequence(s) may then be detected using a hybridization assay which is described in EP 317,077, published May 24, 1989. These hybridization assays, which should detect sequences at the level of 10^6/ml, utilize nucleic acid multimers which bind to single-stranded analyte nucleic acid, and which also bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO 225,807, published June 16, 1987.

The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled and the ingredients for labeling may be included in the kit in separate containers. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, as well as instructions for conducting the test.

Immunoassay and Diagnostic Kits

Both the polypeptides which react immunologically with serum containing HCV antibodies, for example, those detected by the antigenic screening method described infra. in the Examples, as well those derived from or encoded within the isolated clones described in the Examples, and composites thereof, and the antibodies raised against the HCV specific epitopes in these polypeptides, are useful in immunoassays to detect presence of HCV antibodies, or the presence of the virus and/or viral antigens, in biological samples. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. For example, the immunoassay may utilize one viral epitope; alternatively,

the immunoassay may use a combination of viral epitopes derived from these sources; these epitopes may be derived from the same or from different viral polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides. It may use, for example, a monoclonal antibody directed towards a viral epitope(s), a combination of monoclonal antibodies directed towards epitopes of one viral antigen, monoclonal antibodies directed towards epitopes of different viral antigens, polyclonal antibodies directed towards the same viral antigen, or polyclonal antibodies directed towards different viral antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Some of the antigenic regions of the putative polyprotein have been mapped and identified by screening the antigenicitiy of bacterial expression products of HCV cDNAs which encode portions of the polyprotein. See the Examples. Other antigenic regions of HCV may be detected by expressing the portions of the HCV cDNAs in other expression systems, including yeast systems and cellular systems derived from insects and vertebrates. In addition, studies giving rise to an antigenicity index and hydrophobicity/hydrophilicity profile give rise to information concerning the probability of a region's antigenicity.

The studies on antigenic mapping by expression of HCV cDNAs showed that a number of clones containing

these cDNAs expressed polypeptides which were immunologically reactive with serum from individuals with NANBH. No single polypeptide was immunologically reactive with all sera. Five of these polypeptides were very immunogenic in that antibodies to the HCV epitopes in these polypeptides were detected in many different patient sera, although the overlap in detection was not complete. Thus, the results on the immunogenicity of the polypeptides encoded in the various clones suggest that effecient detection systems may include the use of panels of epitopes. The epitopes in the panel may be constructed into one or multiple polypeptides.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing HCV epitopes or antibodies directed against HCV epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay, as well as a suitable set of assay instructions.

Further Characterization of the HCV Genome, Virions, and Viral Antigens Using Probes Derived From cDNA to the Viral Genome

The HCV cDNA sequence information in the newly isolated clones described in the Examples may be used to gain further information on the sequence of the HCV genome, and for identification and isolation of the HCV agent, and thus will aid in its characterization including the nature of the genome, the structure of the viral particle, and the nature of the antigens of which it is composed. This information, in turn, can lead to additional polynucleotide probes, polypeptides derived from the HCV genome, and antibodies directed against HCV

epitopes which would be useful for the diagnosis and/or treatment of HCV caused NANBH.

The cDNA sequence information in the above-mentioned clones is useful for the design of probes for the isolation of additional cDNA sequences which are derived from as yet undefined regions of the HCV genome(s) from which the cDNAs in clones described herein and in EP 0,316,218 are derived. For example, labeled probes containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the 5'-termini or 3'-termini of the composite HCV cDNA sequence shown in Fig. 17 may be used to isolate overlapping cDNA sequences from HCV cDNA libraries. Alternatively, characterization of the genomic segments could be from the viral genome(s) isolated from purified HCV particles. Methods for purifying HCV particles and for detecting them during the purification procedure are described herein, infra. Procedures for isolating polynucleotide genomes from viral particles are known in the art, and one procedure which may be used is that described in EP 0,218,316. The isolated genomic segments could then be cloned and sequenced. An example of this technique, which utilizes amplification of the sequences to be cloned, is provided infra., and yielded clone 16jh.

Methods for constructing cDNA libraries are known in the art, and are discussed supra and infra; a method for the construction of HCV cDNA libraries in lambda-gt11 is discussed in EPO Pub. No. 318,216. However, cDNA libraries which are useful for screening with nucleic acid probes may also be constructed in other vectors known in the art, for example, lambda-gt10 (Huynh et al. (1985)).

Screening for Anti-Viral Agents for HCV

The availability of cell culture and animal model systems for HCV makes it possible to screen for anti-viral agents which inhibit HCV replication, and particularly for those agents which preferentially allow cell growth and multiplication while inhibiting viral replication. These screening methods are known by those of skill in the art. Generally, the anti-viral agents are tested at a variety of concentrations, for their effect on preventing viral replication in cell culture systems which support viral replication, and then for an inhibition of infectivity or of viral pathogenicity (and a low level of toxicity) in an animal model system.

The methods and compositions provided herein for detecting HCV antigens and HCV polynucleotides are useful for screening of anti-viral agents in that they provide an alternative, and perhaps more sensitive means, for detecting the agent's effect on viral replication than the cell plaque assay or ID_{50} assay. For example, the HCV-polynucleotide probes described herein may be used to quantitate the amount of viral nucleic acid produced in a cell culture. This could be accomplished, for example, by hybridization or competition hybridization of the infected cell nucleic acids with a labeled HCV-polynucleotide probe. For example, also, anti-HCV antibodies may be used to identify and quantitate HCV antigen(s) in the cell culture utilizing the immunoassays described herein. In addition, since it may be desirable to quantitate HCV antigens in the infected cell culture by a competition assay, the polypeptides encoded within the HCV cDNAs described herein are useful in these competition assays. Generally, a recombinant HCV polypeptide derived from the HCV cDNA would be labeled, and the inhibition of binding of this labeled polypeptide to an HCV polypeptide due to the antigen produced in the cell culture system would be

monitored. Moreover, these techniques are particularly useful in cases where the HCV may be able to replicate in a cell line without causing cell death.

The anti-viral agents which may be tested for efficacy by these methods are known in the art, and include, for example, those which interact with virion components and/or cellular components which are necessary for the binding and/or replication of the virus. Typical anti-viral agents may include, for example, inhibitors of virion polymerase and/or protease(s) necessary for cleavage of the precursor polypeptides. Other anti-viral agents may include those which act with nucleic acids to prevent viral replication, for example, anti-sense polynucleotides, etc.

Antisense polynucleotides molecules are comprised of a complementary nucleotide sequence which allows them to hybridize specifically to designated regions of genomes or RNAs. Antisense polynucleotides may include, for example, molecules that will block protein translation by binding to mRNA, or may be molecules which prevent replication of viral RNA by transcriptase. They may also include molecules which carry agents (non-covalently attached or covalently bound) which cause the viral RNA to be inactive by causing, for example, scissions in the viral RNA. They may also bind to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability, or chronicity. Antisense molecules which are to hybridize to HCV derived RNAs may be designed based upon the sequence information of the HCV cDNAs provided herein. The antiviral agents based upon anti-sense polynucleotides for HCV may be designed to bind with high specificity, to be of increased solubility, to be stable, and to have low toxicity. Hence, they may be delivered in specialized systems, for example, liposomes, or by gene therapy. In addition, they

may include analogs, attached proteins, substituted or altered bonding between bases, etc.

Other types of drugs may be based upon polynucleotides which "mimic" important control regions of the HCV genome, and which may be therapeutic due to their interactions with key components of the system responsible for viral infectivity or replication.

General Methods

The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, sequencing clones, constructing expression vectors, transforming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, E. coli is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al.

(1977)), the tryptophan (trp) promoter system (Goeddel et al. (1980)) and the lambda-derived P_L promoter and N gene ribosome binding site (Shimatake et al. (1981)) and the hybrid tac promoter (De Boer et al. (1983)) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with E. coli; if desired, other prokaryotic hosts such as strains of Bacillus or Pseudomonas may be used, with corresponding control sequences.

Eukaryotic hosts include yeast and mammalian cells in culture systems. Saccharomyces cerevisiae and Saccharomyces carlsbergensis are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983)), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968); Holland et al. (1978)), including the promoter for 3 phosphoglycerate kinase (Hitzeman (1980)). Terminators may also be included, such as those derived from the enolase gene (Holland (1981)). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are

not naturally associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985, all of which are assigned to the herein assignee, and are hereby incorporated herein by reference.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978)), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.

Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. For example, transformation of the E. coli host cells with lambda-gt11 containing BB-NANBV sequences is discussed in the Example section, infra. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972); Maniatis (1982)). Yeast

transformation by direct uptake may be carried out using the method of Hinnen et al. (1978). Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), or the various known modifications thereof.

Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 microgram of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 microliters buffer solution by incubation of 1-2 hr at 37° C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560.

Sticky ended cleavage fragments may be blunt ended using E. coli DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

Ligation mixtures are transformed into suitable cloning hosts, such as E. coli, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984). If desired the synthetic strands may be labeled with ^{32}P by treatment with polynucleotide kinase in the presence of ^{32}P-ATP, using standard conditions for the reaction.

DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase using, as a primer, a synthetic oligonucleotide complementary to the portion of the DNA to be modified, and having the desired modification included in its own sequence. The resulting double stranded DNA is transformed into a phage supporting host bacterium. Cultures of the transformed bacteria, which contain replications of each strand of the phage, are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and conditions which permit hybridization with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

DNA libraries may be probed using the procedure of Grunstein and Hogness (1975). Briefly, in this procedure, the DNA to be probed is immobilized on nitrocellulose filters, denatured, and prehybridized with a

buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinyl pyrollidone, and Ficoll, 50 mM Na Phosphate (pH 6.5), 0.1% SDS, and 100 micrograms/ml carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42°C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'-^{32}P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

For routine vector constructions, ligation mixtures are transformed into E. coli strain HB101 or other suitable host, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared according to the method of Clewell et al. (1969), usually following chloramphenicol amplification (Clewell (1972)). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing. Sequencing may be by the dideoxy method of Sanger et al. (1977) as further described by Messing et al. (1981), or by the method of Maxam et al. (1980). Problems with band compression, which are sometimes observed in GC rich regions, were

overcome by use of T-deazoguanosine according to Barr et al. (1986).

The enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of anti-body bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

Examples

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present inven-tion. In light of the present disclosure, numerous embodiments within the scope of the claims will be appar-ent to those of ordinary skill in the art.

Isolation and Sequence of Overlapping HCV cDNA Clones 13i, 26j, CA59a, CA84a, CA156e and CA167b

The clones 13i, 26j, CA59a, CA84a, CA156e and CA167b were isolated from the lambda-gt11 library which contains HCV cDNA (ATCC No. 40394), the preparation of which is described in EPO Pub. No. 318,216 (published 31 May 1989), and WO 89/04669 (published 1 June 1989). Screening of the library was with the probes described infra., using the method described in Huynh (1985). The frequencies with which positive clones appeared with the respective probes was about 1 in 50,000.

The isolation of clone 13i was accomplished using a synthetic probe derived from the sequence of clone 12f. The sequence of the probe was:

5' GAA CGT TGC GAT CTG GAA GAC AGG GAC AGG 3'.

The isolation of clone 26j was accomplished using a probe derived from the 5'-region of clone K9-1. The sequence of the probe was:

5' TAT CAG TTA TGC CAA CGG AAG CGG CCC CGA 3'.

The isolation procedures for clone 12f and for clone k9-1 (also called K9-1) are described in EPO Pub. No. 318,216, and their sequences are shown in Figs. 1 and 2, respectively. The HCV cDNA sequences of clones 13i and 26j, are shown in Figs. 4 and 5, respectively. Also shown are the amino acids encoded therein, as well as the overlap of clone 13i with clone 12f, and the overlap of clone 26j with clone 13i. The sequences for these clones confirmed the sequence of clone K9-1. Clone K9-1 had been isolated from a different HCV cDNA library (See EP 0,218,316).

Clone CA59a was isolated utilizing a probe based upon the sequence of the 5'-region of clone 26j. The sequence of this probe was:

5' CTG GTT AGC AGG GCT TTT CTA TCA CCA CAA 3'.

A probe derived from the sequence of clone CA59a was used to isolate clone CA84a. The sequence of the probe used for this isolation was:

5' AAG GTC CTG GTA GTG CTG CTG CTA TTT GCC 3'.

Clone CA156e was isolated using a probe derived from the sequence of clone CA84a. The sequence of the probe was:

5' ACT GGA CGA CGC AAG GTT GCA ATT GCT CTA 3'.

Clone CA167b was isolated using a probe derived from the sequence of clone CA 156e. The sequence of the probe was:

5' TTC GAC GTC ACA TCG ATC TGC TTG TCG GGA 3'.

The nucleotide sequences of the HCV cDNAs in clones CA59a, CA84a, CA156e, and CA167b, are shown Figs. 6, 7, 8, and 9, respectively. The amino acids encoded therein, as well as the overlap with the sequences of relevant clones, are also shown in the Figs.

Creation of "pi" HCV cDNA Library

A library of HCV cDNA, the "pi" library, was constructed from the same batch of infectious chimpanzee plasma used to construct the lambda-gt11 HCV cDNA library (ATCC No. 40394) described in EPO Pub. No. 318,216, and

utilizing essentially the same techniques. However, construction of the pi library utilized a primer-extension method, in which the primer for reverse transcriptase was based on the sequence of clone CA59A. The sequence of the primer was:

5′ GGT GAC GTG GGT TTC 3′.

Isolation and Sequence of Clone pi14a

Screening of the "pi" HCV cDNA library described supra., with the probe used to isolate clone CA167b (See supra.) yielded clone pi14a. The clone contains about 800 base pairs of cDNA which overlaps clones CA167b, CA156e, CA84a and CA59a, which were isolated from the lambda gt-11 HCV cDNA library (ATCC No. 40394). In addition, pi14a also contains about 250 base pairs of DNA which are upstream of the HCV cDNA in clone CA167b.

Isolation and Sequence of Clones CA216a, CA290a and ag30a

Based on the sequence of clone CA167b a synthetic probe was made having the following sequence:

5′ GGC TTT ACC ACG TCA CCA ATG ATT GCC CTA 3′

The above probe was used to screen the , which yielded clone CA216a, whose HCV sequences are shown in Fig. 10.

Another probe was made based on the sequence of clone CA216a having the following sequence:

5′ TTT GGG TAA GGT CAT CGA TAC CCT TAC GTG 3′

Screening the lambda-gt11 library (ATCC No. 40394) with this probe yielded clone CA290a, the HCV sequences therein being shown in Fig. 11.

In a parallel approach, a primer-extension cDNA library was made using nucleic acid extracted from the same infectious plasma used in the original lambda-gt11 cDNA library described above. The primer used was based on the sequence of clones CA216a and CA290a:

5′ GAA GCC GCA CGT AAG 3′

The cDNA library was made using methods similar to those described previously for libraries used in the isolation of clones pi14a and k9-1. The probe used to screen this library was based on the sequence of clone CA290a:

5′ CCG GCG TAG GTC GCG CAA TTT GGG TAA 3′

Clone ag30a was isolated from the new library with the above probe, and contained about 670 basepairs of HCV sequence. See Fig. 12. Part of this sequence overlaps the HCV sequence of clones CA216a and CA290a. About 300 base-pairs of the ag30a sequence, however, is upstream of the sequence from clone CA290a. The non-overlapping sequence shows a start codon (*) and stop codons that may indicate the start of the HCV ORF. Also indicated in Fig. 12 are putative small encoded peptides (#) which may play a role in regulating translation, as well as the putative first amino acid of the putative polypeptide (/), and downstream amino acids encoded therein.

Isolation and Sequence of Clone CA205a

Clone CA205a was isolated from the original lambda gt-11 library (ATCC No. 40394), using a synthetic probe derived from the HCV sequence in clone CA290a (Fig. 11). The sequence of the probe was:

5′ TCA GAT CGT TGG TGG AGT TTA CTT GTT GCC 3′.

The sequence of the HCV cDNA in CA205a, shown in Fig. 13, overlaps with the cDNA sequences in both clones ag30a and CA290a. The overlap of the sequence with that of CA290a is shown by the dotted line above the sequence (the figure also shows the putative amino acids encoded in this fragment).

As observed from the HCV cDNA sequences in clones CA205a and ag30a, the putative HCV polyprotein appears to begin at the ATG start codon; the HCV sequences in both clones contain an in-frame, contiguous double stop codon (TGATAG) forty two nucleotides upstream from this ATG. The HCV ORF appears to begin after these stop codons, and to extend for at least 8907 nucleotides (See the composite HCV cDNA shown in Fig. 17).

Isolation and Sequence of Clone 18g

Based on the sequence of clone ag30a (See Fig. 12) and of an overlapping clone from the original lambda gt-11 library (ATCC No. 40394), CA230a, a synthetic probe was made having the following sequence:

5′ CCA TAG TGG TCT GCG GAA CCG GTG AGT ACA 3′.

Screening of the original lambda-gt11 HCV cDNA library with the probe yielded clone 18g, the HCV cDNA sequence of which is shown in Fig. 14. Also shown in the figure are the overlap with clone ag30a, and putative polypeptides encoded within the HCV cDNA.

The cDNA in clone 18g (C18g or 18g) overlaps that in clones ag30a and CA205a, described supra. The sequence of C18g also contains the double stop codon region observed in clone ag30a. The polynucleotide region upstream of these stop codons presumably represents part of the 5′-region of the HCV genome, which may contain short ORFs, and which can be confirmed by direct sequenc-

ing of the purified HCV genome. These putative small encoded peptides may play a regulatory role in translation. The region of the HCV genome upstream of that represented by C18g can be isolated for sequence analysis using essentially the technique described in EPO Pub. No. 318,216 for isolating cDNA sequences upstream of the HCV cDNA sequence in clone 12f. Essentially, small synthetic oligonucleotide primers of reverse transcriptase, which are based upon the sequence of C18g, are synthesized and used to bind to the corresponding sequence in HCV genomic RNA. The primer sequences are proximal to the known 5'-terminal of C18g, but sufficiently downstream to allow the design of probe sequences upstream of the primer sequences. Known standard methods of priming and cloning ar eused. The resulting cDNA libraries are screened with sequences upstream of the priming sites (as deduced from the elucidated sequence of C18g). The HCV genomic RNA is obtained from either plasma or liver samples from individuals with NANBH. Since HCV appears to be a Flavi-like virus, the 5'-terminus of the genome may be modified with a "cap" structure. It is known that Flavivirus genomes contain 5'-terminal "cap" structures. (Yellow Fever virus, Rice et al. (1988); Dengue virus, Hahn et al (1988); Japanese Encephalitis Virus (1987)).

Isolation and Sequence of Clones from the beta-HCV cDNA library

Clones containing cDNA representative of the 3'-terminal region of the HCV genome were isolated from a cDNA library constructed from the original infectious chimpanzee plasma pool which was used for the creation of the HCV cDNA lambda-gt11 library (ATCC No. 40394), described in EPO Pub. No. 318,216. In order to create the DNA library, RNA extracted from the plasma was "tailed" with poly rA using poly (rA) polymerase, and cDNA was

synthesized using oligo$(dT)_{12-18}$ as a primer for reverse transcriptase. The resulting RNA:cDNA hybrid was digested with RNAase H, and converted to double stranded HCV cDNA. The resulting HCV cDNA was cloned into lambda-gt10, using essentially the technique described in Huynh (1985), yielding the beta (or b) HCV cDNA library. The procedures used were as follows.

An aliquot (12ml) of the plasma was treated with proteinase K, and extracted with an equal volume of phenol saturated with 0.05M Tris-Cl, pH 7.5, 0.05% (v/v) beta-mercaptoethanol, 0.1% (w/v) hydroxyquinolone, 1 mM EDTA. The resulting aqueous phase was re-extracted with the phenol mixture, followed by 3 extractions with a 1:1 mixture containing phenol and chloroform:isoamyl alcohol (24:1), followed by 2 extractions with a mixture of chloroform and isoamyl alcohol (1:1). Subsequent to adjustment of the aqueous phase to 200 mM with respect to NaCl, nucleic acids in the aqueous phase were precipitated overnight at -20°C, with 2.5 volumes of cold absolute ethanol. The precipitates were collected by centrifugation at 10,000 RPM for 40 min., washed with 70% ethanol containing 20 mM NaCl, and with 100% cold ethanol, dried for 5 min. in a dessicator, and dissolved in water.

The isolated nucleic acids from the infectious chimpanzee plasma pool were tailed with poly rA utilizing poly-A polymerase in the presence of human placenta ribonuclease inhibitor (HPRI) (purchased from Amersham Corp.), utilizing MS2 RNA as carrier. Isolated nucleic acids equivalent to that in 2 ml of plasma were incubated in a solution containing TMN (50 mM Tris HCl, pH 7.9, 10 mM $MgCl_2$, 250 mM NaCl, 2.5 mM $MnCl_2$, 2 mM dithiothreitol (DTT)), 40 micromolar alpha-[^{32}P] ATP, 20 units HPRI (Amersham Corp.), and about 9 to 10 units of RNase free poly-A polymerase (BRL). Incubation was for 10 min. at 37°C, and the reactions were stopped with EDTA (final

concentration about 250 mM). The solution was extracted with an equal volume of phenol-chloroform, and with an equal volume of chloroform, and nucleic acids were precipitated overnight at -20°C with 2.5 volumes of ethanol in the presence of 200 mM NaCl.

Isolation of Clone b5a

The beta HCV cDNA library was screened by hybridization using a synthetic probe, which had a sequence based upon the HCV cDNA sequence in clone 15e. The isolation of clone 15e is described in EPO Pub. No. 318,216, and its sequence is shown in Fig. 3. The sequence of the synthetic probe was:

5' ATT GCG AGA TCT ACG GGG CCT GCT ACT CCA 3'.

Screening of the library yielded clone beta-5a (b5a), which contains an HCV cDNA region of approximately 1000 base pairs. The 5'-region of this cDNA overlaps clones 35f, 19g, 26g, and 15e (these clones are described supra). The region between the 3'-terminal poly-A sequence and the 3'-sequence which overlaps clone 15e, contains approximately 200 base pairs. This clone allows the identification of a region of the 3'-terminal sequence the HCV genome.

The sequence of b5a is contained within the sequence of the HCV cDNA in clone 16jh (described infra). Moreover, the sequence is also present in CC34a, isolated from the original lambda-gt11 library (ATCC No. 40394). (The original lambda-gt11 library is referred to herein as the "C" library).

Isolation and Sequence of Clones Generated by PCR Amplification of the 3'-Region of the HCV Genome

Multiple cDNA clones have been generated which contain nucleotide sequences derived from the 3'-region of the HCV genome. This was accomplished by amplifying a targeted region of the genome by a polymerase chain reaction technique described in Saiki et al. (1986), and in Saiki et al. (1988), which was modified as described below. The HCV RNA which was amplified was obtained from the original infectious chimpanzee plasma pool which was used for the creation of the HCV cDNA lambda-gt11 library (ATCC No. 40394) described in EPO Pub. No. 318,216. Isolation of the HCV RNA was as described supra. The isolated RNA was tailed at the 3'-end with ATP by E. coli poly-A polymerase as described in Sippel (1973), except that the nucleic acids isolated from chimp serum were substituted for the nucleic acid substrate. The tailed RNA was then reverse transcribed into cDNA by reverse transcriptase, using an oligo dT-primer adapter, essentially as described by Han (1987), except that the components and sequence of the primer-adapter were:

Stuffer	NotI	SP6 Promoter	Primer
AATTC	GCGGCCGC	CATACGATTTAGGTGACACTATAGAA	T_{15}

The resultant cDNA was subjected to amplification by PCR using two primers:

Primer	Sequence
JH32 (30mer)	ATAGCGGCCGCCCTCGATTGCGAGATCTAC
JH11 (20mer)	AATTCGGGCGGCCGCCATACGA

The JH32 primer contained 20 nucleotide sequences hybridizable to the 5'-end of the target region in the cDNA, with an estimated T_m of 66°C. The JH11 was derived

from a portion of the oligo dT-primer adapter; thus, it is specific to the 3'-end of the cDNA with a T_m of 64°C. Both primers were designed to have a recognition site for the restriction enzyme, NotI, at the 5'-end, for use in subsequent cloning of the amplified HCV cDNA.

The PCR reaction was carried out by suspending the cDNA and the primers in 100 microliters of reaction mixture containing the four deoxynucleoside triphosphates, buffer salts and metal ions, and a thermostable DNA polymerase isolated from Thermus aquaticus (Taq polymerase), which are in a Perkin Elmer Cetus PCR kit (N801-0043 or N801-0055). The PCR reaction was performed for 35 cycles in a Perkin Elmer Cetus DNA thermal cycler. Each cycle consisted of a 1.5 min denaturation step at 94°C, an annealing step at 60°C for 2 min, and a primer extension step at 72°C for 3 min. The PCR products were subjected to Southern blot analysis using a 30 nucleotide probe, JH34, the sequence of which was based upon that of the 3'-terminal region of clone 15e. The sequence of JH34 is:

5' CTT GAT CTA CCT CCA ATC ATT CAA AGA CTC 3'.

The PCR products detected by the HCV cDNA probe ranged in size from about 50 to about 400 base pairs.

In order to clone the amplified HCV cDNA, the PCR products were cleaved with NotI and size selected by polyacrylamide gel electrophoresis. DNA larger than 300 base pairs was cloned into the NotI site of pUC18S The vector pUC18S is constructed by including a NotI polylinker cloned between the EcoRI and SalI sites of pUC18. The clones were screened for HCV cDNA using the JH34 probe. A number of positive clones were obtained and sequenced. The nucleotide sequence of the HCV cDNA insert in one of these clones, 16jh, and the amino acids encoded

therein, are shown in Fig. 15. A nucleotide heterogeneity, detected in the sequence of the HCV cDNA in clone 16jh as compared to another clone of this region, is indicated in the figure.

Compiled HCV cDNA Sequences

An HCV cDNA sequence has been compiled from a series of overlapping clones derived from the various HCV cDNA libraries described supra.. In this sequence, the compiled HCV cDNA sequence obtained from clones b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, and CA59a is upstream of the compiled HCV cDNA sequence published in EPO Pub. No. 318,216, which is shown in Fig. 16. The compiled HCV cDNA sequence obtained from clones b5a and 16jh downstream of the compiled HCV cDNA sequence published in EPO Pub. No. 318,216.

Fig. 17 shows the compiled HCV cDNA sequence derived from the above-described clones and the compiled HCV cDNA sequence published in EPO Pub. No. 318,216. The clones from which the sequence was derived are b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, CA59a, K9-1 (also called k9-1),26j, 13i, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a, and 16jh. In the figure the three dashes above the sequence indicate the position of the putative initiator methionine codon.

Clone b114a was obtained using the cloning procedure described for clone b5a, supra., except that the probe was the synthetic probe used to detect clone 18g, supra. Clone b114a overlaps with clones 18g, ag30a, and CA205a, except that clone b114a contains an extra two nucleotides upstream of the sequence in clone 18g (i.e., 5'-CA). These extra two nucleotides have been included in the HCV genomic sequence shown in Fig. 17.

It should be noted that although several of the clones described supra. have been obtained from libraries other than the original HCV cDNA lambda-gt11 C library (ATCC No. 40394), these clones contain HCV cDNA sequences which overlap HCV cDNA sequences in the original library. Thus, essentially all of the HCV sequence is derivable from the original lambda-gt11 C library (ATCC No. 40394) which was used to isolate the first HCV cDNA clone (5-1-1). The isolation of clone 5-1-1 is described in EPO Pub. No. 318,216.

Purification of Fusion Polypeptide C100-3 (Alternate method)

The fusion polypeptide, C100-3 (also called HCV c100-3 and alternatively, c100-3), is comprised of superoxide dismutase (SOD) at the N-terminus an in-frame C100 HCV polypeptide at the C-terminus. A method for preparing the polypeptide by expression in yeast, and differential extraction of the insoluble fraction of the extracted host yeast cells, is described in EPO Pub. No. 318,216. An alternative method for the preparation of this fusion polypeptide is described below. In this method the antigen is precipitated from the crude cell lysate with acetone; the acetone precipitated antigen is then subjected to ion-exchange chromatography, and further purified by gel filtration.

The fusion polypeptide, C100-3 (HCV c100-3), is expressed in yeast strain JSC 308 (ATCC No. 20879) transformed with pAB24C100-3 (ATCC No. 67976); the transformed yeast are grown under conditions which allow expression (i.e., by growth in YEP containing 1% glucose). (See EPO Pub. No. 318,216). A cell lysate is prepared by suspending the cells in Buffer A (20 mM Tris HCl, pH 8.0, 1 mM EDTA, 1 mM PMSF. The cells are broken by grinding with glass beads in a Dynomill type homogenizer or its

equivalent. The extent of cell breakage is monitored by counting cells under a microscope with phase optics. Broken cells appear dark, while viable cells are light-colored. The percentage of broken cells is determined.

When the percentage of broken cells is approximately 90% or greater, the broken cell debris is separated from the glass beads by centrifugation, and the glass beads are washed with Buffer A. After combining the washes and homogenate, the insoluble material in the lysate is obtained by centrifugation. The material in the pellet is washed to remove soluble proteins by suspension in Buffer B (50 mM glycine, pH 12.0, 1 mM DTT, 500 mM NaCl), followed by Buffer C (50 mM glycine, pH 10.0, 1 mM DTT). The insoluble material is recovered by centrifugation, and solubilized by suspension in Buffer C containing SDS. The extract solution may be heated in the presence of beta-mercaptoethanol and concentrated by ultrafiltration. The HCV c100-3 in the extract is precipitated with cold acetone. If desired, the precipitate may be stored at temperatures at about or below -15°C.

Prior to ion exchange chromatography, the acetone precipitated material is recovered by centrifugation, and may be dried under nitrogen. The precipitate is suspended in Buffer D (50 mM glycine, pH 10.0, 1 mM DTT, 7 M urea), and centrifuged to pellet insoluble material. The supernatant material is applied to an anion exchange column previously equilibrated with Buffer D. Fractions are collected and analyzed by ultraviolet absorbance or gel electrophoresis on SDS polyacrylamide gels. Those fractions containing the HCV c100-3 polypeptide are pooled.

In order to purify the HCV c100-3 polypeptide by gel filtration, the pooled fractions from the ion-exchange column are heated in the presence of beta-mercaptoethanol

and SDS, and the eluate is concentrated by ultrafiltration. The concentrate is applied to a gel filtration column previously equilibrated with Buffer E (20 mM Tris HCl, pH 7.0, 1 mM DTT, 0.1% SDS). The presence of HCV c100-3 in the eluted fractions, as well as the presence of impurities, are determined by gel electrophoresis on polyacrylamide gels in the presence of SDS and visualization of the polypeptides. Those fractions containing purified HCV c100-3 are pooled. Fractions high in HCV c100-3 may be further purified by repeating the gel filtration process. If the removal of particulate material is desired, the HCV c100-3 containing material may be filtered through a 0.22 micron filter.

Expression and Antigenicity of Polypeptides Encoded in HCV cDNA

Polypeptides Expressed in E. coli

The polypeptides encoded in a number of HCV cDNAs which span the HCV genomic ORF were expressed in E. coli, and tested for their antigenicity using serum obtained from a variety of individuals with NANBH. The expression vectors containing the cloned HCV cDNAs were constructed from pSODcf1 (Steimer et al. (1986). In order to be certain that a correct reading frame would be achieved, three separate expression vectors, pcf1AB, pcf1CD, and pcf1EF were created by ligating either of three linkers, AB, CD, and EF to a BamHI-EcoRI fragment derived by digesting to completion the vector pSODcf1 with EcoRI and BamHI, followed by treatment with alkaline phosphatase. The linkers were created from six oligomers, A, B, C, D, E, and F. Each oligomer was phosphorylated by treatment with kinase in the presence of ATP prior to annealing to its complementary oligomer. The sequences of the synthetic linkers were the following.

Name	DNA Sequence (5′ to 3′)
A	GATC CTG AAT TCC TGA TAA
B	GAC TTA AGG ACT ATT TTA A
C	GATC CGA ATT CTG TGA TAA
D	GCT TAA GAC ACT ATT TTA A
E	GATC CTG GAA TTC TGA TAA
F	GAC CTT AAG ACT ATT TTA A

Each of the three linkers destroys the original EcoRI site, and creates a new EcoRI site within the linker, but within a different reading frame. Hence, the HCV cDNA EcoRI fragments isolated from the clones when inserted into the expression vector, were in three different reading frames.

The HCV cDNA fragments in the designated lambda-gt11 clones were excised by digestion with EcoRI; each fragment was inserted into pcf1AB, pcf1CD, and pcf1EF. These expression constructs were then transformed into D1210 E. coli cells, the transformants were cloned, and recombinant bacteria from each clone were induced to express the fusion polypeptides by growing the bacteria in the presence of IPTG.

Expression products of the indicated HCV cDNAs were tested for antigenicity by direct immunological screening of the colonies, using a modification of the method described in Helfman et al. (1983). Briefly, as shown in Fig. 18, the bacteria were plated onto nitrocellulose filters overlaid on ampicillin plates to give approximately 1,000 colonies per filter. Colonies were replica plated onto nitrocellulose filters, and the replicas were regrown overnight in the presence of 2 mM

IPTG and ampicillin. The bacterial colonies were lysed by suspending the nitrocellulose filters for about 15 to 20 min in an atmosphere saturated with $CHCl_3$ vapor. Each filter then was placed in an individual 100 mm Petri dish containing 10 ml of 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 3% (w/v) BSA, 40 micrograms/ml lysozyme, and 0.1 microgram/ml DNase. The plates were agitated gently for at least 8 hours at room temperature. The filters were rinsed in TBST (50 mM Tris HCl, pH8.0, 150 mM NaCl, 0.005% Tween 20). After incubation, the cell residues were rinsed and incubated in TBS (TBST without Tween) containing 10% sheep serum; incubation was for 1 hour. The filters were then incubated with pretreated sera in TBS from individuals with NANBH, which included: 3 chimpanzees; 8 patients with chronic NANBH whose sera were positive with respect to antibodies to HCV C100-3 polypeptide (described in EPO Pub. No. 318,216, and supra.) (also called C100); 8 patients with chronic NANBH whose sera were negative for anti-C100 antibodies; a convalescent patient whose serum was negative for anti-C100 antibodies; and 6 patients with community acquired NANBH, including one whose sera was strongly positive with respect to anti-C100 antibodies, and one whose sera was marginally positive with respect to anti-C100 antibodies. The sera, diluted in TBS, was pretreated by preabsorption with hSOD. Incubation of the filters with the sera was for at least two hours. After incubation, the filters were washed two times for 30 min with TBST. Labeling of expressed proteins to which antibodies in the sera bound was accomplished by incubation for 2 hours with ^{125}I-labeled sheep anti-human antibody. After washing, the filters were washed twice for 30 min with TBST, dried, and autoradiographed.

A number of clones (see infra.) expressed polypeptides containing HCV epitopes which were im-

munologically reactive with serum from individuals with NANBH. Five of these polypeptides were very immunogenic in that antibodies to HCV epitopes in these polypeptides were detected in many different patient sera. The clones encoding these polypeptides, and the location of the polypeptide in the putative HCV polyprotein (wherein the amino acid numbers begin with the putative initiator codon) are the following: clone 5-1-1, amino acids 1694-1735; clone C100, amino acids 1569-1931; clone 33c, amino acids 1192-1457; clone CA279a, amino acids 1-84; and clone CA290a amino acids 9-177. The location of the immunogenic polypeptides within the putative HCV polyprotein are shown immediately below.

Clones encoding polypeptides of proven reactivity with sera from NANBH patients.

Clone	Location within the HCV polyprotein (amino acid no. beginning with putative initiator methionine)
CA279a	1-84
CA74a	437-582
13i	511-690
CA290a	9-177
33c	1192-1457
40b	1266-1428
5-1-1	1694-1735
81	1689-1805
33b	1916-2021
25c	1949-2124
14c	2054-2223
8f	2200-3325
33f	2287-2385
33g	2348-2464
39c	2371-2502
15e	2796-2886
C100	1569-1931

The results on the immunogenicity of the polypeptides encoded in the various clones examined suggest efficient detection and immunization systems may include panels of HCV polypeptides/epitopes.

Expression of HCV Epitopes in Yeast

Three different yeast expression vectors which allow the insertion of HCV cDNA into three different read-

ing frames are constructed. The construction of one of the vectors, pAB24C100-3 is described in EPO Pub. No. 318,216. In the studies below, the HCV cDNA from the clones listed in supra. in the antigenicity mapping study using the E. coli expressed products are substituted for the C100 HCV cDNA. The construction of the other vectors replaces the adaptor described in the above E. coli studies with one of the following adaptors:

Adaptor 1

ATT TTG AAT TCC TAA TGA G
AC TTA AGG ATT ACT CAG CT

Adaptor 2

AAT TTG GAA TTC TAA TGA G
AC CTT AAG ATT ACT CAG CT.

The inserted HCV cDNA is expressed in yeast transformed with the vectors, using the expression conditions described supra. for the expression of the fusion polypeptide, C100-3. The resulting polypeptides are screened using the sera from individuals with NANBH, described supra. for the screening of immunogenic polypeptides encoded in HCV cDNAs expressed in E. coli.

Comparison of the Hydrophobic Profiles of HCV Polyproteins with West Nile Virus Polyprotein and with Dengue Virus NS1

The hydrophobicity profile of an HCV polyprotein segment was compared with that of a typical Flavivirus, West Nile virus. The polypeptide sequence of the West Nile virus polyprotein was deduced from the known polynucleotide sequences encoding the non-structural

proteins of that virus. The HCV polyprotein sequence was deduced from the sequence of overlapping cDNA clones. The profiles were determined using an antigen program which uses a window of 7 amino acid width (the amino acid in question, and 3 residues on each side) to report the average hydrophobicity about a given amino acid residue. The parameters giving the reactive hydrophobicity for each amino acid residue are from Kyte and Doolittle (1982). Fig. 19 shows the hydrophobic profiles of the two polyproteins; the areas corresponding to the non-structural proteins of West Nile virus, ns1 through ns5, are indicated in the figure. As seen in the figure, there is a general similarity in the profiles of the HCV polyprotein and the West Nile virus polyprotein.

The sequence of the amino acids encoded in the 5'-region of HCV cDNA shown in Fig. 16 has been compared with the corresponding region of one of the strains of Dengue virus, described supra., with respect to the profile of regions of hydrophobicity and hydrophilicity (data not shown). This comparison indicated that the polypeptides from HCV and Dengue encoded in this region, which corresponds to the region encoding NS1 (or a portion thereof), have a similar hydrophobic/hydrophilic profile.

The similarity in hydrophobicity profiles, in combination with the previously identified homologies in the amino acid sequences of HCV and Dengue Flavivirus in EP 0,218,316 suggests that HCV is related to these members of the Flavivirus family.

Characterization of the Putative Polypeptides Encoded Within the HCV ORF

The sequence of the HCV cDNA sense strand, shown in Fig. 17, was deduced from the overlapping HCV cDNAs in the various clones described in EPO Pub. No. 318,216 and those described supra. It may be deduced from the

sequence that the HCV genome contains primarily one long continuous ORF, which encodes a polyprotein. In the sequence, nucleotide number 1 corresponds to the first nucleotide of the initiator MET codon; minus numbers indicate that the nucleotides are that distance away in the 5'-direction (upstream), while positive numbers indicate that the nucleotides are that distance away in the 3'-direction (downstream). The composite sequence shows the "sense" strand of the HCV cDNA.

The amino acid sequence of the putative HCV polyprotein deduced from the HCV cDNA sense strand sequence is also shown in Fig. 17, where position 1 begins with the putative initiator methionine.

Possible protein domains of the encoded HCV polyprotein, as well as the approximate boundaries, are the following (the polypeptides identified within the parentheses are those which are encoded in the Flavivirus domain):

Putative Domain	Approximate Boundary (amino acid nos.)
"C" (nucleocapsid protein)	1-120
"E" (Virion envelope protein(s) and possibly matrix (M) proteins	120-400
"NS1" (complement fixation antigen?)	400-660
"NS2" (unknown function)	660-1050
"NS3" (protease?)	1050-1640

"NS4" (unknown function)	1640-2000
"NS5" (polymerase)	2000-? end

It should be noted, however, that hydrophobicity profiles (described infra), indicate that HCV diverges from the Flavivirus model, particularly with respect to the region upstream of NS2. Moreover, the boundaries indicated are not intended to show firm demarcations between the putative polypeptides.

The Hydrophilic and Antigenic Profile of the Polypeptide

Profiles of the hydrophilicity/hydrophobicity and the antigenic index of the putative polyprotein encoded in the HCV cDNA sequence shown in Fig. 16 were determined by computer analysis. The program for hydrophilicity/hydrophobicity was as described supra. The antigenic index results from a computer program which relies on the following criteria: 1)surface probability, 2) prediction of alpha-helicity by two different methods; 3) prediction of beta-sheet regions by two different methods; 4) prediction of U-turns by two different methods; 5) hydrophilicity/hydrophobicity; and flexibility. The traces of the profiles generated by the computer analyses are shown in Fig. 20. In the hydrophilicity profile, deflection above the abscissa indicates hydrophilicity, and below the abscissa indicates hydrophobicity. The probability that a polypeptide region is antigenic is usually considered to increase when there

is a deflection upward from the abscissa in the hydrophilic and/or antigenic profile. It should be noted, however, that these profiles are not necessarily indicators of the strength of the immunogenicity of a polypeptide.

Identification of Co-linear Peptides in HCV and Flaviviruses

The amino acid sequence of the putative polyprotein encoded in the HCV cDNA sense strand was compared with the known amino acid sequences of several members of Flaviviruses. The comparison shows that homology is slight, but due to the regions in which it is found, it is probably significant. The conserved co-linear regions are shown in Fig. 21. The amino acid numbers listed below the sequences represent the number in the putative HCV polyprotein (See Fig. 17.)

The spacing of these conserved motifs is similar between the Flaviviruses and HCV, and implies that there is some similarity between HCV and these flaviviral agents.

The following listed materials are on deposit under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

lambda-gt11	ATCC No.	Deposit Date
HCV cDNA library	40394	1 Dec. 1987
clone 81	40388	17 Nov. 1987
clone 91	40389	17 Nov. 1987
clone 1-2	40390	17 Nov. 1987
clone 5-1-1	40391	18 Nov. 1987
clone 12f	40514	10 Nov. 1988
clone 35f	40511	10 Nov. 1988
clone 15e	40513	10 Nov. 1988
clone K9-1	40512	10 Nov. 1988
JSC 308	20879	5 May 1988
pS356	67683	29 April 1988

In addition, the following deposits were made on 11 May 1989.

Strain		Linkers	ATCC No.
D1210	(Cf1/5-1-1)	EF	67967
D1210	(Cf1/81)	EF	67968
D1210	(Cf1/CA74a)	EF	67969
D1210	(Cf1/35f)	AB	67970
D1210	(Cf1/279a)	EF	67971
D1210	(Cf1/C36)	CD	67972
D1210	(Cf1/13i)	AB	67973
D1210	(Cf1/C33b)	EF	67974
D1210	(Cf1/CA290a)	AB	67975
HB101	(AB24/C100 #3R)		67976

The following derivatives of strain D1210 were deposited on 3 May 1989.

Strain Derivative	ATCC No.
pCF1CS/C8f	67956
pCF1AB/C12f	67952
pCF1EF/14c	67949
pCF1EF/15e	67954
pCF1AB/C25c	67958
pCF1EF/C33c	67953
pCF1EF/C33f	67050
pCF1CD/33g	67951
pCF1CD/C39c	67955
pCF1EF/C40b	67957
pCF1EF/CA167b	67959

The following strains were deposited on May 12, 1989.

Strain	ATCC No.
Lambda gt11(C35)	40603
Lambda gt10(beta-5a)	40602
D1210 (C40b)	67980
D1210 (M16)	67981

Upon allowance and issuance of this application as a United States Patent, all restriction on availability of these deposits will be irrevocably removed; and access to the designated deposits will be available during pendency of the above-named application to one determined by the Commissioner to be entitled thereto under 37 CFR 1.14 and 35 USC 1.22. Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. The deposited materials mentioned herein are intended for convenience only, and are not required to practice the present invention in view

of the descriptions herein, and in addition these materials are incorporated herein by reference.

Industrial Applicability

The invention, in the various manifestations disclosed herein, has many industrial uses, some of which are the following. The HCV cDNAs may be used for the design of probes for the detection of HCV nucleic acids in samples. The probes derived from the cDNAs may be used to detect HCV nucleic acids in, for example, chemical synthetic reactions. They may also be used in screening programs for anti-viral agents, to determine the effect of the agents in inhibiting viral replication in cell culture systems, and animal model systems. The HCV polynucleotide probes are also useful in detecting viral nucleic acids in humans, and thus, may serve as a basis for diagnosis of HCV infections in humans.

In addition to the above, the cDNAs provided herein provide information and a means for synthesizing polypeptides containing epitopes of HCV. These polypeptides are useful in detecting antibodies to HCV antigens. A series of immunoassays for HCV infection, based on recombinant polypeptides containing HCV epitopes are described herein, and will find commercial use in diagnosing HCV induced NANBH, in screening blood bank donors for HCV-caused infectious hepatitis, and also for detecting contaminated blood from infectious blood donors. The viral antigens will also have utility in monitoring the efficacy of anti-viral agents in animal model systems. In addition, the polypeptides derived from the HCV cDNAs disclosed herein will have utility as vaccines for treatment of HCV infections.

The polypeptides derived from the HCV cDNAs, besides the above stated uses, are also useful for raising anti-HCV antibodies. Thus, they may be used in anti-HCV

vaccines. However, the antibodies produced as a result of immunization with the HCV polypeptides are also useful in detecting the presence of viral antigens in samples. Thus, they may be used to assay the production of HCV polypeptides in chemical systems. The anti-HCV antibodies may also be used to monitor the efficacy of anti-viral agents in screening programs where these agents are tested in tissue culture systems. They may also be used for passive immunotherapy, and to diagnose HCV caused NANBH by allowing the detection of viral antigen(s) in both blood donors and recipients. Another important use for anti-HCV antibodies is in affinity chromatography for the purification of virus and viral polypeptides. The purified virus and viral polypeptide preparations may be used in vaccines. However, the purified virus may also be useful for the development of cell culture systems in which HCV replicates.

Antisense polynucleotides may be used as inhibitors of viral replication.

For convenience, the anti-HCV antibodies and HCV polypeptides, whether natural or recombinant, may be packaged into kits.

CLAIMS

1. A recombinant polynucleotide comprising a sequence derived from HCV cDNA, wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

2. A recombinant polynucleotide according to claim 1, encoding an epitope of HCV.

3. A recombinant vector comprising the polynucleotide of claim 1 or claim 2.

4. A host cell transformed with the vector of claim 3.

5. A recombinant expression system comprising an open reading frame (ORF) of DNA derived from the recombinant polynucleotide of claim 1 or claim 2, wherein the ORF is operably linked to a control sequence compatible with a desired host.

6. A cell transformed with the recombinant expression system of claim 5.

7. A polypeptide produced by the cell of claim 6.

8. A purified polypeptide comprising an epitope encoded within HCV cDNA wherein the HCV cDNA is of a

sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

9. An immunogenic polypeptide produced by a cell transformed with a recombinant expression vector comprising an ORF of DNA derived from HCV cDNA, wherein the HCV cDNA is comprised of a sequence derived from the HCV cDNA sequence in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the ORF is operably linked to a control sequence compatible with a desired host.

10. A peptide comprising an HCV epitope, wherein the peptide is of the formula

$$AA_x\text{-}AA_y,$$

wherein x and y designate amino acid numbers shown in Fig. 17, and wherein the peptide is selected from the group consisting of AA1-AA25, AA1-AA50, AA1-AA84, AA9-AA177, AA1-AA10, AA5-AA20, AA20-AA25, AA35-AA45, AA50-AA100, AA40-AA90, AA45-AA65, AA65-AA75, AA80-90, AA99-AA120, AA95-AA110, AA105-AA120, AA100-AA150, AA150-AA200, AA155-AA170, AA190-AA210, AA200-AA250, AA220-AA240, AA245-AA265, AA250-AA300, AA290-AA330, AA290-305, AA300-AA350, AA310-AA330, AA350-AA400, AA380-AA395, AA405-AA495, AA400-AA450, AA405-AA415, AA415-AA425, AA425-AA435, AA437-AA582, AA450-AA500, AA440-AA460, AA460-AA470, AA475-AA495, AA500-AA550, AA511-AA690, AA515-AA550, AA550-AA600, AA550-AA625, AA575-AA605, AA585-AA600, AA600-AA650, AA600-AA625, AA635-AA665, AA650-AA700, AA645-AA680, AA700-AA750, AA700-AA725, AA700-AA750, AA725-AA775, AA770-AA790, AA750-AA800, AA800-AA815,

AA825-AA850, AA850-AA875, AA800-AA850, AA920-AA990, AA850-AA900, AA920-AA945, AA940-AA965, AA970-AA990, AA950-AA1000, AA1000-AA1060, AA1000-AA1025, AA1000-AA1050, AA1025-AA1040, AA1040-AA1055, AA1075-AA1175, AA1050-AA1200, AA1070-AA1100, AA1100-AA1130, AA1140-AA1165, AA1192-AA1457, AA1195-AA1250, AA1200-AA1225, AA1225-AA1250, AA1250-AA1300, AA1260-AA1310, AA1260-AA1280, AA1266-AA1428, AA1300-AA1350, AA1290-AA1310, AA1310-AA1340, AA1345-AA1405, AA1345-AA1365, AA1350-AA1400, AA1365-AA1380, AA1380-AA1405, AA1400-AA1450, AA1450-AA1500, AA1460-AA1475, AA1475-AA1515, AA1475-AA1500, AA1500-AA1550, AA1500-AA1515, AA1515-AA1550, AA1550-AA1600, AA1545-AA1560, AA1569-AA1931, AA1570-AA1590, AA1595-AA1610, AA1590-AA1650, AA1610-AA1645, AA1650-AA1690, AA1685-AA1770, AA1689-AA1805, AA1690-AA1720, AA1694-AA1735, AA1720-AA1745, AA1745-AA1770, AA1750-AA1800, AA1775-AA1810, AA1795-AA1850, AA1850-AA1900, AA1900-AA1950, AA1900-AA1920, AA1916-AA2021, AA1920-AA1940, AA1949-AA2124, AA1950-AA2000, AA1950-AA1985, AA1980-AA2000, AA2000-AA2050, AA2005-AA2025, AA2020-AA2045, AA2045-AA2100, AA2045-AA2070, AA2054-AA2223, AA2070-AA2100, AA2100-AA2150, AA2150-AA2200, AA2200-AA2250, AA2200-AA2325, AA2250-AA2330, AA2255-AA2270, AA2265-AA2280, AA2280-AA2290, AA2287-AA2385, AA2300-AA2350, AA2290-AA2310, AA2310-AA2330, AA2330-AA2350, AA2350-AA2400, AA2348-AA2464, AA2345-AA2415, AA2345-AA2375, AA2370-AA2410, AA2371-AA2502, AA2400-AA2450, AA2400-AA2425, AA2415-AA2450, AA2445-AA2500, AA2445-AA2475, AA2470-AA2490, AA2500-AA2550, AA2505-AA2540, AA2535-AA2560, AA2550-AA2600, AA2560-AA2580, AA2600-AA2650, AA2605-AA2620, AA2620-AA2650,

AA2640-AA2660, AA2650-AA2700, AA2655-AA2670, AA2670-AA2700, AA2700-AA2750, AA2740-AA2760, AA2750-AA2800, AA2755-AA2780, AA2780-AA2830, AA2785-AA2810, AA2796-AA2886, AA2810-AA2825, AA2800-AA2850, AA2850-AA2900, AA2850-AA2865, AA2885-AA2905, AA2900-AA2950, AA2910-AA2930, AA2925-AA2950, AA2945-end(C' terminal).

11. A polypeptide comprised of the peptide of claim 10.

12. An immunogenic polypeptide attached to a solid substrate, wherein the polypeptide is according to claim 7, or claim 8, or claim 9, or claim 10, or claim 11, or wherein the polypeptide is comprised of an epitope encoded within HCV cDNA wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

13. A monoclonal antibody directed against an epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

14. A preparation of purified polyclonal antibodies directed against a polypeptide comprised of an epitope encoded within HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone

CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

15. A polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or from the complement of the HCV cDNA sequence.

16. A kit for analyzing samples for the presence of polynucleotides from HCV comprising a polynucleotide probe containing a nucleotide sequence of about 8 or more nucleotides, wherein the nucleotide sequence is derived from HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, wherein the polynucleotide probe is in a suitable container.

17. A kit for analyzing samples for the presence of an HCV antigen comprising an antibody which reacts immunologically with an HCV antigen, wherein the antigen contains an epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

18. A kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide containing an HCV epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or

clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

19. A kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide expressed from HCV cDNA in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, wherein the antigenic polypeptide is present in a suitable container.

20. A method for detecting HCV nucleic acids in a sample comprising:

(a) reacting nucleic acids of the sample with a polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, and wherein the reacting is under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample,

(b) detecting a polynucleotide duplex which contains the probe, formed in step (a).

21. An immunoassay for detecting an HCV antigen comprising:

(a) incubating a sample suspected of containing an HCV antigen with an antibody directed against an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone

CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and (b) detecting an antibody-antigen complex formed in step (a) which contains the antibody.

22. An immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing anti-HCV antibodies with an antigen polypeptide containing an epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and

(b) detecting an antibody-antigen complex formed in step (a) which contains the antigen polypeptide.

23. An immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing anti-HCV antibodies with the polypeptide of claim 9, under conditions which allow formation of an antigen-antibody complex; and

(b) detecting an antibody-antigen complex formed in step (a) which contains the antigen polypeptide.

24. A vaccine for treatment of HCV infection comprising an immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or

8659 to 8866 in Fig. 17 or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

25. A method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is of the sequence present in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

26. An antisense polynucleotide derived from HCV cDNA, wherein the HCV cDNA is that shown in Fig. 17.

27. A method for preparing purified fusion polypeptide C100-3 comprising:

(a) providing a crude cell lysate containing polypeptide C100-3,

(b) treating the crude cell lysate with an amount of acetone which causes the polypeptide to precipitate,

(c) isolating and solubilizing the precipitated material,

(d) isolating the C100-3 polypeptide by anion exchange chromatography, and

(e) further isolating the C100-3 polypeptide of step (d) by gel filtration.

28. A method for preparing an HCV polypeptide comprising:

(a) providing a host cell transformed with a recombinant expression system comprising an open reading frame (ORF) of DNA derived from HCV cDNA, wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, wherein the ORF is operably linked to a control sequence compatible with a desired host; and

(b) incubating the host cell under conditions with allow expression of the HCV polypeptide.

29. A method for preparing an immunogenic HCV polypeptide comprising:

(a) providing a host cell transformed with a recombinant expression vector comprising an ORF of DNA derived from HCV cDNA, wherein the HCV cDNA is comprised of a sequence derived from the HCV cDNA sequence in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33c, or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, wherein the OEF is operably linked to a control sequence compatible with the desired host; and

(b) incubating the host cell under conditions which allow expression of the HCV polypeptide.

30. A method for preparing a host cell

transformed with a recombinant polynucleotide comprising a sequence of HCV cDNA derived from the HCV cDNA in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17 comprising:

(a) providing a host cell capable of transformation;

(b) providing the recombinant polynucleotide; and

(c) incubating (a) with (b) under conditions which allow transformation of the host cell with the polynucleotide.

31. A method for preparing a recombinant polynucleotide comprised of a sequence of HCV cDNA derived from the HCV cDNA in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17 comprising:

(a) providing a host cell transformed with the recombinant polynucleotide; and

(b) isolating said polynucleotide from said host cell.

32. A method for preparing blood free of HCV comprising:

(a) providing a sample of blood suspected of containing HCV and anti-HCV antibodies;

(b) providing an immunogenic polypeptide prepared according to claim 28 or 29;

(c) incubating the sample of (a) with the immunogenic polypeptide of (b) under conditions which allow the formation of antibody-HCV polypeptide complexes;

(d) detecting the complexes formed in step (c); and

(e) saving the blood from which complexes were not detected in (d).

33. A method for preparing blood free of HCV comprising:

(a) providing nucleic acids from a sample of blood suspected of containing HCV polynucleotides;

(b) providing a probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17,

(c) reacting (a) with (b) under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample;

(d) detecting a polynucleotide which contains the probe, formed in step (c); and

(e) saving the blood from which complexes were not detected in (d).

34. A method for producing a hybridoma which produces anti-HCV monoclonal antibodies comprising:

(a) immunizing an individual with an immunogenic polypeptide containing an epitope encoded in HCV cDNA, wherein the HCV cDNA is HCV cDNA in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone

CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17; or

(b) immunizing an individual with an immunogenic polypeptide prepared according to claim 29;

(c) immortalizing antibody producing cells from the immunized individual;

(d) selecting an immortal cell which produces antibodies which react with an HCV epitope in the immunogenic polypeptide of (a) or (b); and

(e) growing said immortal cell.

FIG. 1 Translation of DNA 12f

```
     IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn
1    CCATATTTAAAATCAGGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCA
     GGTATAAATTTTAGTCCTACATGCACCCTCCCCAGCTTGTGTCCGACCTTCGACGGACGT

     TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu
61   ACTGGACGCGGGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGT
     TGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCA

     LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu
121  TACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCT
     ATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGATGGTCGGA

     SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal
181  TGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGG
     ACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCC

                                                  ----------
     GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeu
241  TGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCTTC
     ACCCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCTCATGCAGCAAGAGGACAAGGAAG

     ------------------------------------------------------------
     LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu
301  TGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGG
     ACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCC

     AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu
361  AGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTC
     TCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAG

     -------
     Val
421  TTGTATC
     AACATAG
```

FIG. 2-1 Translation of DNA k9-1

```
      GlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGly
1     CAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGG
      GTCCGACAGGACTCTCCGATCGGTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCC

      ProIleSerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrPro
61    GCCCTATCAGTTATGCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACC
      CGGGATAGTCAATACGGTTGCCTTCGCCGGGGCTGGTCGCGGGGATGACGACCGTGATGG

      ProLysProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThr
121   CCCCAAAACCTTGCGGTATTGTGCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCA
      GGGGTTTTGGAACGCCATAACACGGGCGCTTCTCACACACACCAGGCCATATAACGAAGT

      ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
181   CTCCCAGCCCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGG
      GAGGGTCGGGGCACCACCACCCTTGCTGGCTGTCCAGCCCGCGCGGGTGGATGTCGACCC

      GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
241   GTGAAAATGATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGT
      CACTTTTACTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCA

      GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
301   TCGGTTGTACCTGGATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTG
      AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACAC

      IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
361   TCATCGGAGGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATC
      AGTAGCCTCCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAG

      AspAlaThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAsp
421   CGGACGCCACATACTCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCG
      GCCTGCGGTGTATGAGAGCCACGCCGAGGCCAGGGACCTAGTGTGGGTCCACGGACCAGC

                                                    ----------
      TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
481   ACTACCCGTATAGGCTTTGGCATTATCCTTGTACCATCAACTACACTATATTTAAAATCA
      TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGATATAAATTTTAGT

      ----------
      MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
541   GGATGTACGTGGGAGGGGTCGAGCACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCG
      CCTACATGCACCCTCCCCAGCTCGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGC

      ----------
      ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
601   AACGTTGCGATCTGGAAGATAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTA
      TTGCAACGCTAGACCTTCTATCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGAT

      ----------
      GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIle
661   CACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTGCCAGCCTTGTCCACCGGCCTCA
      GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGACGGTCGGAACAGGTGGCCGGAGT

      HisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAla
721   TCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCG
      AGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCCACCCCAGTTCGTAGC

      ----------
      SerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArg
781   CGTCCTGGGCCATTAAGTGGGAGTACGTCGTCCTCCTGTTCCTTCTGCTTGCAGACGCGC
      GCAGGACCCGGTAATTCACCCTCATGCAGCAGGAGGACAAGGAAGACGAACGTCTGCGCG

      ----------
```

```
                ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsn
841   GCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAAGCGGCTTTGGAGA
      CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCCTTCGCCGAAACCTCT

                ---------------------------------------------------------
                LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal
901   ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCG
      TGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGC

                ---------------------------------------------------------
                PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe
961   TGTTCTTCTGCTTTGCATGGTATCTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCT
      ACAAGAAGACGAAACGTACCATAGACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGA

                ---------------------------------------------------------
                TyrGlyMetTrpProLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu
1021  TCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGC
      AGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCG

                ---------------------------------------------------------
                AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr
1081  TGGACACGGAGGTGGCCGCGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTAA
      ACCTGTGCCTCCACCGGCGCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGATT

                ---------------------------------------------------------
                LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu
1141  CTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTC
      GAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAG

                ---------------------------------------------------------
                ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg
1201  TGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGC
      ACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCG

                ---------------------------------------------------------
                AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys
1261  GCGACGCTGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCA
      CGCTGCGACAGTAGAATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGT

                ---------------------------------------------
                LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAla
1321  AATTGCTGCTGGCCGTCTTCGGACCCCTTTGGATTCTTCAAGCCAG
      TTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGTC
```

FIG. 2-2

FIG. 3

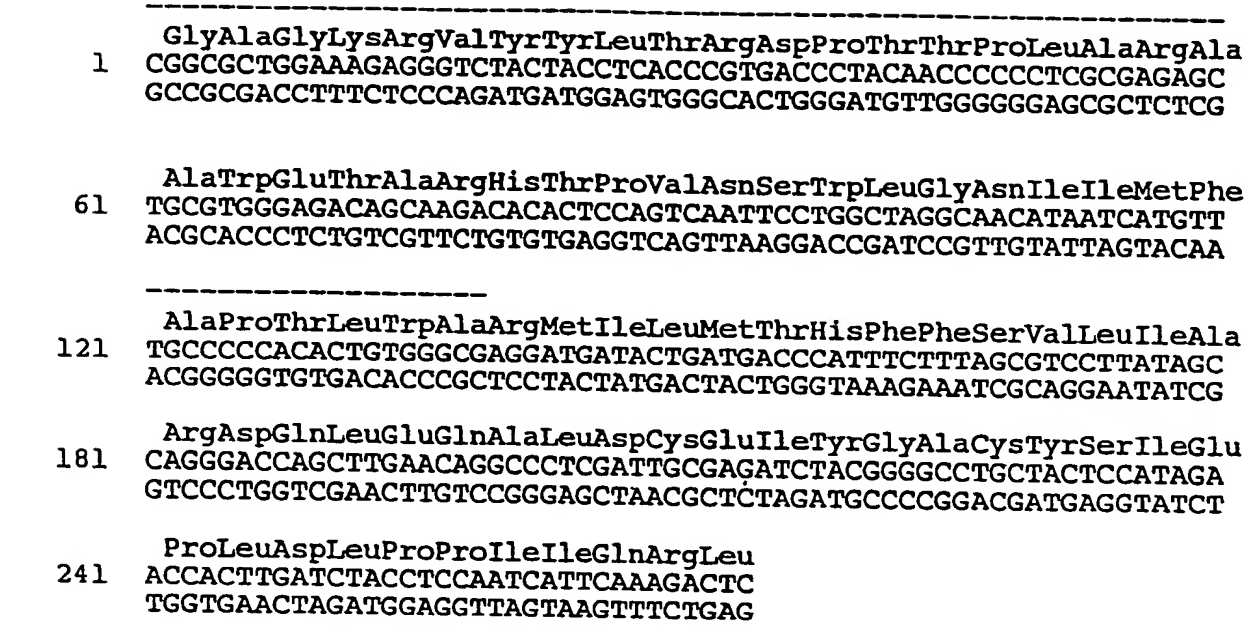


FIG. 5

Translation of DNA 26j

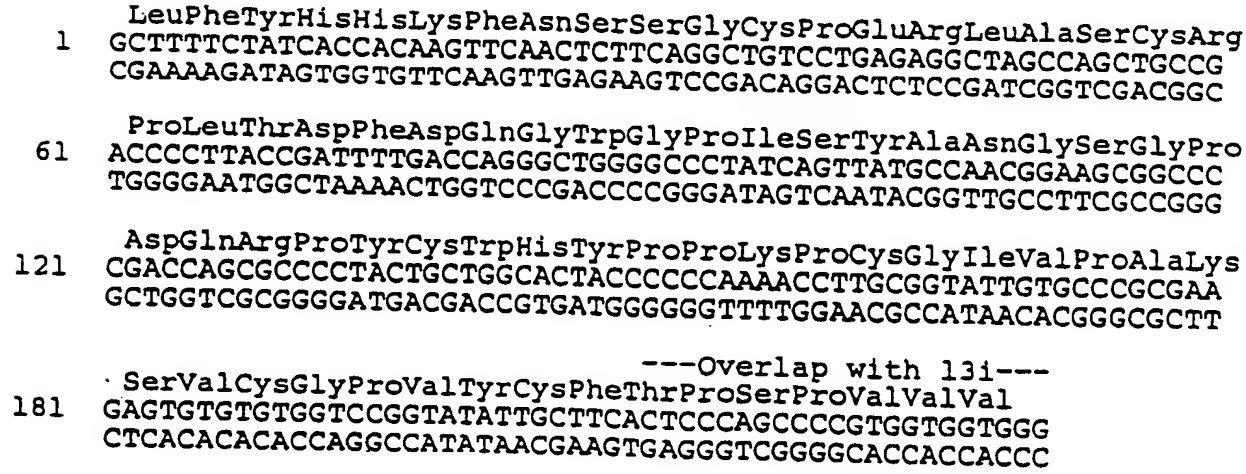


FIG. 4

Translation of DNA 13i

```
      ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
1     CTCCCAGCCCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCCTACCTACAGCTGGG
      GAGGGTCGGGGCACCACCACCCTTGCTGGCTGTCCAGCCCGCGCGGATGGATGTCGACCC

      GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
61    GTGAAAATGATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGT
      CACTTTTACTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCA

      GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
121   TCGGTTGTACCTGGATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTG
      AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACAC

      IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
181   TCATCGGAGGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATC
      AGTAGCCTCCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAG

      AspAlaThrTyrSerArgCysGlySerGlyProTrpLeuThrProArgCysLeuValAsp
241   CGGACGCCACATACTCTCGGTGCGGCTCCGGTCCCTGGCTCACACCCAGGTGCCTGGTCG
      GCCTGCGGTGTATGAGAGCCACGCCGAGGCCAGGGACCGAGTGTGGGTCCACGGACCAGC

                                                    ----------
      TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
301   ACTACCCGTATAGGCTTTGGCATTATCCTTGTACCATCAACTACACCATATTTAAAATCA
      TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAATTTTAGT

      ------------------------------------------------------------
      MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
361   GGATGTACGTGGGAGGGGTCGAGCACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCG
      CCTACATGCACCCTCCCCAGCTCGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGC

      ---------------Overlap with 12f----------------------------
      ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
421   AACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTA
      TTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGAT

      ------------------------------------------------------------
      GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeu
481   CACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTGCCAGCCTTGTCCACCGGCCTCA
      GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTTGGGACGGTCGGAACAGGTGGCCGGAGT
```

FIG. 6

Translation of DNA CA59a

```
     LeuValMetAlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIleAlaGlyAla
  1  TTGGTAATGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCT
     AACCATTACCGAGTCGACGAGGCCTAGGGTGTTCGGTAGAACCTGTACTAGCGACCACGA

     HisTrpGlyValLeuAlaGlyIleAlaTyrPheSerMetValGlyAsnTrpAlaLysVal
 61  CACTGGGGAGTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTC
     GTGACCCCTCAGGACCGCCCGTATCGCATAAAGAGGTACCACCCCTTGACCCGCTTCCAG

     LeuValValLeuLeuLeuPheAlaGlyValAspAlaGluThrHisValThrGlyGlySer
121  CTGGTAGTGCTGCTGCTATTTGCCGGCGTCGACGCGGAAACCCACGTCACCGGGGGAAGT
     GACCATCACGACGACGATAAACGGCCGCAGCTGCGCCTTTGGGTGCAGTGGCCCCCTTCA

     AlaGlyHisThrValSerGlyPheValSerLeuLeuAlaProGlyAlaLysGlnAsnVal
181  GCCGGCCACACTGTGTCTGGATTTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTC
     CGGCCGGTGTGACACAGACCTAAACAATCGGAGGAGCGTGGTCCGCGGTTCGTCTTGCAG

     GlnLeuIleAsnThrAsnGlySerTrpHisLeuAsnSerThrAlaLeuAsnCysAsnAsp
241  CAGCTGATCAACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAACTGCAATGAT
     GTCGACTAGTTGTGGTTGCCGTCAACCGTGGAGTTATCGTGCCGGGACTTGACGTTACTA

                                                              -----
     SerLeuAsnThrGlyTrpLeuAlaGlyLeuPheTyrHisHisLysPheAsnSerSerGly
301  AGCCTCAACACCGGCTGGTTGGCAGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGC
     TCGGAGTTGTGGCCGACCAACCGTCCCGAAAAGATAGTGGTGTTCAAGTTGAGAAGTCCG
                                  ------Overlap with 26j------------

     ------Overlap with K9-1--------
     CysProGluArgLeuAlaSerCysArgPro
361  TGTCCTGAGAGGCTAGCCAGCTGCCGACCCC
     ACAGGACTCTCCGATCGGTCGACGGCTGGGG
     --------------------------------
```

Translation of DNA CA84a FIG. 7

```
      GlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMetAlaTrpAsp
  1   CGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCGCATGGCATGGG
      GCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGCCCAGTGGCGTACCGTACCC

                                                  ----------
      MetMetMetAsnTrpSerProThrThrAlaLeuValMetAlaGlnLeuLeuArgIlePro
 61   ATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAATGGCTCAGCTGCTCCGGATCC
      TATACTACTACTTGACCAGGGGATGCTGCCGCAACCATTACCGAGTCGACGAGGCCTAGG

      ------------------------------------------------------------
      GlnAlaIleLeuAspMetIleAlaGlyAlaHisTrpGlyValLeuAlaGlyIleAlaTyr
121   CACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCGT
      GTGTTCGGTAGAACCTGTACTAGCGACCACGAGTGACCCCTCAGGACCGCCCGTATCGCA

      ------------Overlap with CA59a------------------------------
      PheSerMetValGlyAsnTrpAlaLysValLeuValValLeuLeuLeuPheAlaGlyVal
181   ATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGCG
      TAAAGAGGTACCACCCCTTGACCCGCTTCCAGGACCATCACGACGACGATAAACGGCCGC

      ----------------------------
      AspAlaGluThrHisValThrGly
241   TCGACGCGGAAACCCACGTCACCGGGG
      AGCTGCGCCTTTGGGTGCAGTGGCCCC
```

Translation of DNA CA156e FIG. 8

```
      CysTrpValAlaMetThrProThrValAlaThrArgAspGlyLysLeuProAlaThrGln
  1   GTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGACGCA
      CACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGGCGCTGCGT

      LeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCysSerAlaLeuTyrVal
 61   GCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCCCTCTACGT
      CGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACAAGCCGGGAGATGCA

      GlyAspLeuCysGlySerValPheLeuValGlyGlnLeuPheThrPheSerProArgArg
121   GGGGGACCTATGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCG
      CCCCCTGGATACGCCCAGACAGAAAGAACAGCCGGTTGACAAGTGGAAGAGAGGGTCCGC

                  ------------------------------------------------
      HisTrpThrThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArg
181   CCACTGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCG
      GGTGACCTGCTGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGCCCAGTGGC

      --------------------Overlap with CA84a----------------------
      MetAlaTrpAspMetMetMetAsnTrpSerProThrThrAlaLeuValValAlaGlnLeu
241   CATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAGTGGCTCAGCT
      GTACCGTACCCTATACTACTACTTGACCAGGGGATGCTGCCGCAACCATCACCGAGTCGA

      --------------------
      LeuArgIleProGlnAla
301   GCTCCGGATCCCACAAGCC
      CGAGGCCTAGGGTGTTCGG
```

FIG. 9

Translation of DNA CA167b

```
      SerThrGlyLeuTyrHisValThrAsnAspCysProAsnSerSerIleValTyrGluAla
  1   CTCCACGGGGCTTTACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGC
      GAGGTGCCCCGAAATGGTGCAGTGGTTACTAACGGGATTGAGCTCATAACACATGCTCCG

      AlaAspAlaIleLeuHisThrProGlyCysValProCysValArgGluGlyAsnAlaSer
 61   GGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTC
      CCGGCTACGGTAGGACGTGTGAGGCCCCACGCAGGGAACGCAAGCACTCCCGTTGCGGAG

        ------------------------------------------------------------
      ArgCysTrpValAlaMetThrProThrValAlaThrArgAspGlyLysLeuProAlaThr
121   GAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGAC
      CTCCACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGGCGCTG

      ------------------Overlap with CA156e----------------------
      GlnLeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCysSerAlaLeuTyr
181   GCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCTACCCTCTGTTCGGCCCTCTA
      CGTCGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGATGGGAGACAAGCCGGGAGAT

      ---------------------------------
      ValGlyAspLeuCysGlySerValPheLeu
241   CGTGGGGGACTTGTGCGGGTCTGTCTTTCTTG
      GCACCCCCTGAACACGCCCAGACAGAAAGAAC
```

FIG. 10

Translation of DNA ssCA216a

FIG. 11

Translation of DNA ssCA290a

```
      LysLysAsnLysArgAsnThrAsnArgArgProGlnAspValLysPheProGlyGlyGly
1     AAAAAAAAAACAAACGTAACACCAACCGTCGCCCACAGGACGTCAAGTTCCCGGGTGGCG
      TTTTTTTTTTGTTTGCATTGTGGTTGGCAGCGGGTGTCCTGCAGTTCAAGGGCCCACCGC

      GlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeuGlyValArgAla
61    GTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTAGATTGGGTGTGCGCG
      CAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAACCCACACGCGC

      ThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnProIleProLysAla
121   CGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGACGCCAGCCTATCCCCAAGG
      GCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCGGTCGGATAGGGGTTCC

      ArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsn
181   CTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCA
      GAGCAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCCATGGGAACCGGGGAGATACCGT

      GluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGly
241   ATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGGCCTAGCTGGG
      TACTCCCGACGCCCACCCGCCCTACCGAGGACAGAGGGGCACCGAGAGCCGGATCGACCC

                  ----------------------------------------------------
      ProThrAspProArgArgArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCys
301   GCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGATACCCTTACGT
      CGGGGTGTCTGGGGGCCGCATCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCA

      ----------------------------------------------------------------
      GlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeuGlyGlyAlaAla
361   GCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCCCTCTTGGAGGCGCTG
      CGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAGCCGCGGGGAGAACCTCCGCGAC

      ------------------------overlap with CA216a-------------------
      ArgAlaLeuAlaHisGlyValArgValLeuGluAspGlyValAsnTyrAlaThrGlyAsn
421   CCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAACTATGCAACAGGGA
      GGTCCCGGGACCGCGTACCGCAGGCCCAAGACCTTCTGCCGCACTTGATACGTTGTCCCT

      ----------------------------
      LeuProGlyCysSerPheSerThrPhe
481   ACCTTCCTGGTTGCTCTTTCTCTACCTTC
      TGGAAGGACCAACGAGAAAGAGATGGAAG
```

Translation of DNA ag30a

FIG. 12-1

```
                                                #MetSerValValGlnProProGlyProProLeu
                                #MetAlaLeuValOP
  1  CGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCC
     GCGTCTTTCGCAGATCGGTACCGCAATCATACTCACAGCACGTCGGAGGTCCTGGGGGGG

       ProGlyGluProAM

 61  TCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGAC
     AGGGCCCTCTCGGTATCACCAGACGCCTTGGCCACTCATGTGGCCTTAACGGTCCTGCTG

                                                #MetProGlyAspLeuGlyValProProGlnAsp
121  CGGGTCCTTTCTTGGATCAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCAAGA
     GCCCAGGAAAGAACCTAGTTGGGCGAGTTACGGACCTCTAAACCCGCACGGGGGCGTTCT

                                                              OP AM GlyAlaCys
      CysAM                                                             *
181  CTGCTAGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTT
     GACGATCGGCTCATCACAACCCAGCGCTTTCCGGAACACCATGACGGACTATCCCACGAA

                                                        |
       GluCysProGlyArgSerArgArgProCysThrMetSerThrAsnProLysProGlnLys

                                                                    -
241  GCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAA
     CGCTCACGGGGCCCTCCAGAGCATCTGGCACGTGGTACTCGTGCTTAGGATTTGGAGTTT

       LysAsnLysArgAsnThrAsnArgArgProGlnAspValLysPheProGlyGlyGlyGln
     ------------------------------------------------------------
301  AAAAAAACAAACGTAACACCAACCGTCGCCCACAGGACGTCAAGTTCCCGGGTGGCGGTC
     TTTTTTTGTTTGCATTGTGGTTGGCAGCGGGTGTCCTGCAGTTCAAGGGCCCACCGCCAG

       IleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeuGlyValArgAlaThr
     ------------------------------------------------------------
361  AGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGA
     TCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAACCCACACGCGCGCT

       ArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnProIleProLysAlaArg
     ------------------------------------------------------------
421  CGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGACGTCAGCCTATCCCCAAGGCTC
     GCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGATAGGGGTTCCGAG

       ArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGlu
     ----------------------overlap with CA290a-----------------
481  GTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAATG
     CAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCCATGGGAACCGGGGAGATACCGTTAC

       GlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyPro
     ------------------------------------------------------------
541  AGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGGCCTAGCTGGGGCC
     TCCCGACGCCCACCCGCCCTACCGAGGACAGAGGGGCACCGAGAGCCGGATCGACCCCGG

       ThrAspProArgArgArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCysGly
```

```
601  CCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCG
     GGTGTCTGGGGGCCGCATCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCACGC

       Phe

     -----
661  GCTTC
     CGAAG
```

* = Start of long HCV ORF
| = Putative first amino acid of large HCV polyprotein
= Putative small encoded peptides (that may play a translational regulatory role)

FIG. 12-2

FIG. 13

Translation of DNA CA205a

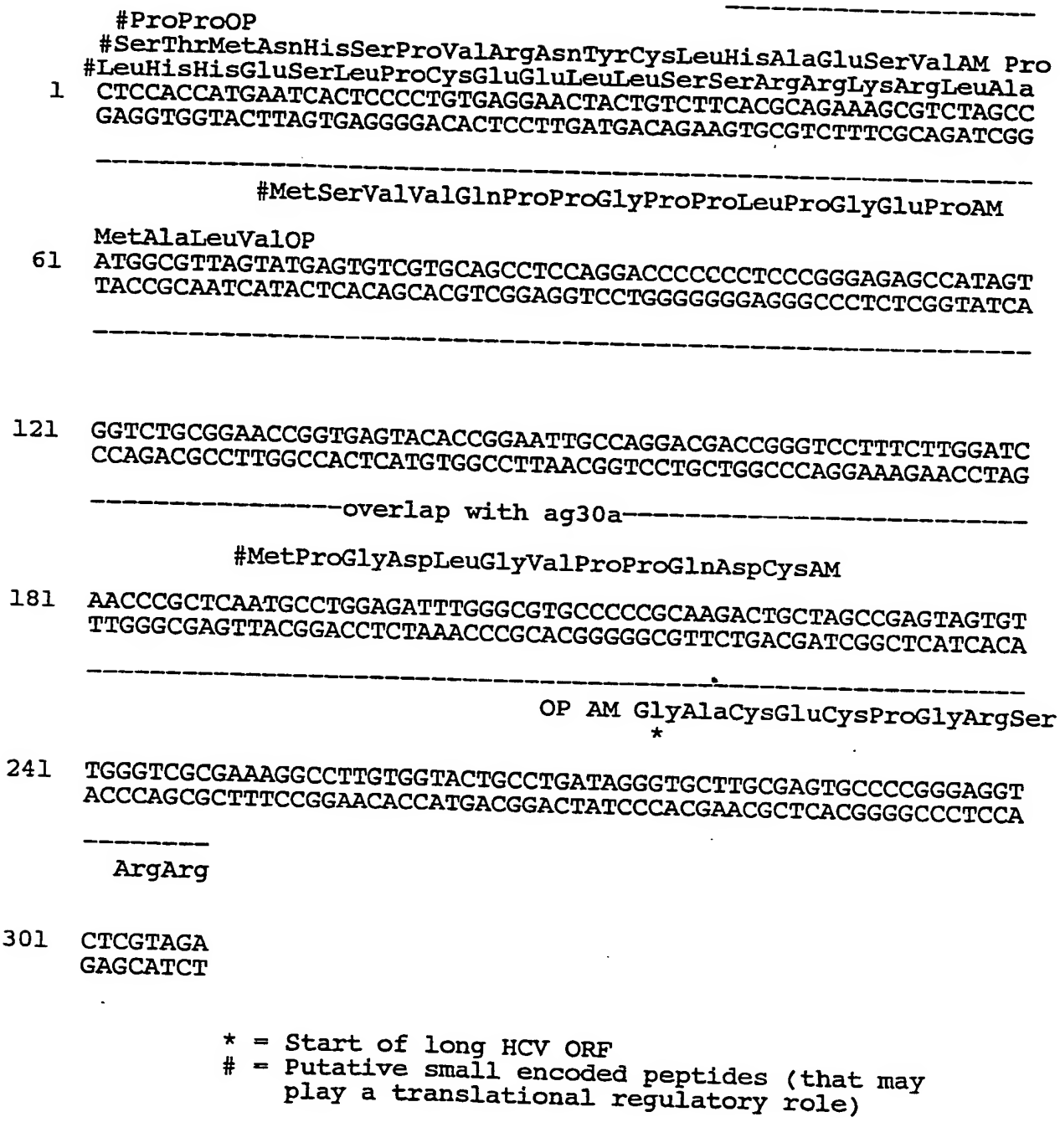
```
     ValLeuGlyArgGluArgProCysGlyThrAlaOP AM GlyAlaCysGluCysProGly
1    GTCTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGG
     CAGAACCCAGCGCTTTCCGGAACACCATGACGGACTATCCCACGAACGCTCACGGGGCCC

                  *                                  ----------
     ArgSerArgArgProCysThrMetSerThrAsnProLysProGlnArgLysThrLysArg
61   AGGTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGT
     TCCAGAGCATCTGGCACGTGGTACTCGTGCTTAGGATTTGGAGTTTCTTTTTGGTTTGCA
     ------------------------------------------------------------
     AsnThrAsnArgArgProGlnAspValLysPheProGlyGlyGlyGlnIleValGlyGly
121  AACACCAACCGTCGCCCACAGGACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGA
     TTGTGGTTGGCAGCGGGTGTCCTGCAGTTCAAGGGCCCACCGCCAGTCTAGCAACCACCT
     ------------------------------------------------------------
     ValTyrLeuLeuProArgArgGlyProArgLeuGlyValArgAlaThrArgLysThrSer
181  GTTTACTTGTTGCCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCC
     CAAATGAACAACGGCGCGTCCCCGGGATCTAACCCACACGCGCGCTGCTCTTTCTGAAGG
     ----------overlap with CA290a----------
     GluArgSerGlnProArgGlyArgArgGlnProIleProLysAlaArgArgProGluGly
241  GAGCGGTCGCAACCTCGAGGTAGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGC
     CTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGATAGGGGTTCCGAGCAGCCGGGCTCCCG
     ------------------------------------------------------------
     ArgThrTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys
301  AGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAATGAGGGCTGCG
     TCCTGGACCCGAGTCGGGCCCATGGGAACCGGGGAGATACCGTTACTCCCGACGC
```

* = putative initiator methionine codon

FIG. 14

Translation of DNA 18g



* = Start of long HCV ORF
\# = Putative small encoded peptides (that may play a translational regulatory role)

FIG. 15

Translation of DNA 16jh

```
      ------Overlap with 15e ----------------------------
      GlyAlaCysTyrSerIleGluProLeuAspLeuProProIleIleGlnArgLeuHisGly
   1  GGGGCCTGCTACTCCATAGAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGC
      CCCCGGACGATGAGGTATCTTGGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCG

      LeuSerAlaPheSerLeuHisSerTyrSerProGlyGluIleAsnArgValAlaAlaCys
  61  CTCAGCGCATTTTCACTCCACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGC
      GAGTCGCGTAAAAGTGAGGTGTCAATGAGAGGTCCACTTTAATTATCCCACCGGCGTACG

                                                 Gly*
                                                 G
      LeuArgLysLeuGlyValProProLeuArgAlaTrpArgHisArgAlaArgSerValArg
 121  CTCAGAAAACTTGGGGTACCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGC
      GAGTCTTTTGAACCCCATGGCGGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCG

      AlaArgLeuLeuAlaArgGlyGlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrp
 181  GCTAGGCTTCTGGCCAGAGGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGG
      CGATCCGAAGACCGGTCTCCTCCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACC

      AlaValArgThrLysLeuLys
 241  GCAGTAAGAACAAAGCTCAAAC
      CGTCATTCTTGTTTCGAGTTTG
```

* = nucleotide heterogeneity

COMBINED ORF OF DNAs pi14a THROUGH 15e

FIG. 16-1

(pi14a/CA167b/CA156e/CA84a/CA59a/K9-1/12f/14i/11b/7f/7e/
8h/33c/40b/37b/35/36/81/32/33b/25c/14c/8f/33f/33g/39c/
35f/19g/26g & 15e)

```
                ArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMet
    1   AGGTCGCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATG
        TCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCACGCCGAAGCGGCTGGAGTAC

                GlyTyrIleProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGly
   61   GGGTACATACCGCTCGTCGGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGC
        CCCATGTATGGCGAGCAGCCGCGGGGAGAACCTCCGCGACGGTCCCGGGACCGCGTACCG

                ValArgValLeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPhe
  121   GTCCGGGTTCTGGAAGACGGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTC
        CAGGCCCAAGACCTTCTGCCGCACTTGATACGTTGTCCCTTGGAAGGACCAACGAGAAAG

                SerIlePheLeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGlnVal
  181   TCTATCTTCCTTCTGGCCCTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGTG
        AGATAGAAGGAAGACCGGGACGAGAGAACGAACTGACACGGGCGAAGCCGGATGGTTCAC

                ArgAsnSerThrGlyLeuTyrHisValThrAsnAspCysProAsnSerSerIleValTyr
  241   CGCAACTCCACGGGGCTTTACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTAC
        GCGTTGAGGTGCCCCGAAATGGTGCAGTGGTTACTAACGGGATTGAGCTCATAACACATG

                GluAlaAlaAspAlaIleLeuHisThrProGlyCysValProCysValArgGluGlyAsn
  301   GAGGCGGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAAC
        CTCCGCCGGCTACGGTAGGACGTGTGAGGCCCCACGCAGGGAACGCAAGCACTCCCGTTG

                AlaSerArgCysTrpValAlaMetThrProThrValAlaThrArgAspGlyLysLeuPro
  361   GCCTCGAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCC
        CGGAGCTCCACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGG

                AlaThrGlnLeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCysSerAla
  421   GCGACGCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCC
        CGCTGCGTCGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACAAGCCGG

                LeuTyrValGlyAspLeuCysGlySerValPheLeuValGlyGlnLeuPheThrPheSer
  481   CTCTACGTGGGGGACCTATGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCT
        GAGATGCACCCCCTGGATACGCCCAGACAGAAAGAACAGCCGGTTGACAAGTGGAAGAGA

                ProArgArgHisTrpThrThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThr
  541   CCCAGGCGCCACTGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACG
        GGGTCCGCGGTGACCTGCTGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGC

                GlyHisArgMetAlaTrpAspMetMetMetAsnTrpSerProThrThrAlaLeuValMet
  601   GGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAATG
        CCAGTGGCGTACCGTACCCTATACTACTACTTGACCAGGGGATGCTGCCGCAACCATTAC

                AlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIleAlaGlyAlaHisTrpGly
  661   GCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGA
        CGAGTCGACGAGGCCTAGGGTGTTCGGTAGAACCTGTACTAGCGACCACGAGTGACCCCT

                ValLeuAlaGlyIleAlaTyrPheSerMetValGlyAsnTrpAlaLysValLeuValVal
  721   GTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTG
        CAGGACCGCCCGTATCGCATAAAGAGGTACCACCCCTTGACCCGCTTCCAGGACCATCAC

                LeuLeuLeuPheAlaGlyValAspAlaGluThrHisValThrGlyGlySerAlaGlyHis
  781   CTGCTGCTATTTGCCGGCGTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCAC
        GACGACGATAAACGGCCGCAGCTGCGCCTTTGGGTGCAGTGGCCCCCTTCACGGCCGGTG
```

FIG. 16-2

```
                ThrValSerGlyPheValSerLeuLeuAlaProGlyAlaLysGlnAsnValGlnLeuIle
841             ACTGTGTCTGGATTTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGATC
                TGACACAGACCTAAACAATCGGAGGAGCGTGGTCCGCGGTTCGTCTTGCAGGTCGACTAG

                AsnThrAsnGlySerTrpHisLeuAsnSerThrAlaLeuAsnCysAsnAspSerLeuAsn
901             AACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAAC
                TTGTGGTTGCCGTCAACCGTGGAGTTATCGTGCCGGGACTTGACGTTACTATCGGAGTTG

                ThrGlyTrpLeuAlaGlyLeuPheTyrHisHisLysPheAsnSerSerGlyCysProGlu
961             ACCGGCTGGTTGGCAGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCCTGAG
                TGGCCGACCAACCGTCCCGAAAAGATAGTGGTGTTCAAGTTGAGAAGTCCGACAGGACTC

                ArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGlyProIleSerTyr
1021            AGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTAT
                TCCGATCGGTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCCCGGGATAGTCAATA

                AlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrProProLysProCys
1081            GCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAACCTTGC
                CGGTTGCCTTCGCCGGGGCTGGTCGCGGGGATGACGACCGTGATGGGGGGTTTTGGAACG

                GlyIleValProAlaLysSerValCysGlyProValTyrCysPheThrProSerProVal
1141            GGTATTGTGCCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGTG
                CCATAACACGGGCGCTTCTCACACACACCAGGCCATATAACGAAGTGAGGGTCGGGGCAC

                ValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGlyGluAsnAspThr
1201            GTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGGGTGAAAATGATACG
                CACCACCCTTGCTGGCTGTCCAGCCCGCGCGGGTGGATGTCGACCCCACTTTTACTATGC

                AspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPheGlyCysThrTrp
1261            GACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGG
                CTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCAAGCCAACATGGACC

                MetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysValIleGlyGlyAla
1321            ATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGGGCG
                TACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACACAGTAGCCTCCCCGC

                GlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisProAspAlaThrTyr
1381            GGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATAC
                CCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAGGCCTGCGGTGTATG

                SerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAspTyrProTyrArg
1441            TCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCGACTACCCGTATAGG
                AGAGCCACGCCGAGGCCAGGGACCTAGTGTGGGTCCACGGACCAGCTGATGGGCATATCC

                LeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArgMetTyrValGly
1501            CTTTGGCATTATCCTTGTACCATCAACTACACCATATTTAAAATCAGGATGTACGTGGGA
                GAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAATTTTAGTCCTACATGCACCCT

                GlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGluArgCysAspLeu
1561            GGGGTCGAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGCGATCTG
                CCCCAGCTTGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGCTTGCAACGCTAGAC

                GluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThrGlnTrpGlnVal
1621            GAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTC
                CTTCTGTCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGATGTGTCACCGTCCAG

                LeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIleHisLeuHisGln
1681            CTCCCGTGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAG
                GAGGGCACAAGGAAGTGTTGGGATGGTCGGAACAGGTGGCCGGAGTAGGTGGAGGTGGTC

                AsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAlaSerTrpAlaIle
1741            AACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCATT
                TTGTAACACCTGCACGTCATGAACATGCCCCACCCCAGTTCGTAGCGCAGGACCCGGTAA

                LysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCys
1801            AAGTGGGAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGC
```

FIG. 16-3

```
                TTCACCCTCATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCGCGCAGACGAGGACG

                LeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeu
1861            TTGTGGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTT
                AACACCTACTACGATGAGTATAGGGTTCGCCTCCGCCGAAACCTCTTGGAGCATTATGAA

                AsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePheCysPhe
1921            AATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTT
                TTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAA

                AlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrp
1981            GCATGGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGG
                CGTACCATAAACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCCTACACC

                ProLeuLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluVal
2041            CCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTG
                GGAGAGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCGACCTGTGCCTCCAC

                AlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyr
2101            GCCGCGTCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATAT
                CGGCGCAGCACACCGCCACAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATA

                TyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGlu
2161            TACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAA
                ATGTTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTT

                AlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArgAspAlaValIle
2221            GCGCAACTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATC
                CGCGTTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCCGCGCTGCGGCAGTAG

                LeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLysLeuLeuLeuAla
2281            TTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCC
                AATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGTTTAACGACGACCGG

                ValPheGlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValProTyrPheValArg
2341            GTCTTCGGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGC
                CAGAAGCCTGGGGAAACCTAAGAAGTTCGGTCAAACGAATTTCATGGGATGAAACACGCG

                ValGlnGlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrVal
2401            GTCCAAGGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTG
                CAGGTTCCGGAAGAGGCCAAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCAC

                GlnMetValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThr
2461            CAAATGGTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACT
                GTTTACCAGTAGTAATTCAATCCCCGCGAATGACCGTGGATACAAATATTGGTAGAGTGA

                ProLeuArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAlaValGluProVal
2521            CCTCTTCGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTC
                GGAGAAGCCCTGACCCGCGTGTTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTCAG

                ValPheSerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGly
2581            GTCTTCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGT
                CAGAAGAGGGTTTACCTCTGGTTCGAGTAGTGCACCCCCCGTCTATGGCGGCGCACGCCA

                AspIleIleAsnGlyLeuProValSerAlaArgArgGlyArgGluIleLeuLeuGlyPro
2641            GACATCATCAACGGCTTGCCTGTTTCCGCCCGCAGGGGCCGGGAGATACTGCTCGGGCCA
                CTGTAGTAGTTGCCGAACGGACAAAGGCGGGCGTCCCCGGCCCTCTATGACGAGCCCGGT

                AlaAspGlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGln
2701            GCCGATGGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAG
                CGGCTACCTTACCAGAGGTTCCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCGGGTC

                GlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGln
2761            CAGACAAGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAA
                GTCTGTTCCCCGGAGGATCCCACGTATTAGTGGTCGGATTGACCGGCCCTGTTTTTGGTT
```

FIG. 16-4

```
                ValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIle
2821            GTGGAGGGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATC
                CACCTCCCACTCCAGGTCTAACACAGTTGACGACGGGTTTGGAAGGACCGTTGCACGTAG

                AsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys
2881            AATGGGGTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAG
                TTACCCCACACGACCTGACAGATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTC

                GlyProValIleGlnMetTyrThrAsnValAspGlnAspLeuValGlyTrpProAlaPro
2941            GGTCCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCG
                CCAGGACAGTAGGTCTACATATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGC

                GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThr
3001            CAAGGTAGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACG
                GTTCCATCGGCGAGTAACTGTGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGC

                ArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSer
3061            AGGCACGCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCG
                TCCGTGCGGCTACAGTAAGGGCACGCGGCCGCCCCACTATCGTCCCCGTCGGACGACAGC

                ProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGly
3121            CCCCGGCCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGG
                GGGGCCGGGTAAAGGATGAACTTTCCGAGGAGCCCCCCAGGCGACAACACGGGGCGCCCC

                HisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaValAsp
3181            CACGCCGTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGAC
                GTGCGGCACCCGTATAAATCCCGGCGCCACACGTGGGCACCTCACCGATTCCGCCACCTG

                PheIleProValGluAsnLeuGluThrThrMetArgSerProValPheThrAspAsnSer
3241            TTTATCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCC
                AAATAGGGACACCTCTTGGATCTCTGTTGGTACTCCAGGGGCCACAAGTGCCTATTGAGG

                SerProProValValProGlnSerPheGlnValAlaHisLeuHisAlaProThrGlySer
3301            TCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGC
                AGAGGTGGTCATCACGGGGTCTCGAAGGTCCACCGAGTGGAGGTACGAGGGTGTCCGTCG

                GlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeu
3361            GGCAAAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTC
                CCGTTTTCGTGGTTCCAGGGCCGACGTATACGTCGAGTCCCGATATTCCACGATCATGAG

                AsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIle
3421            AACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATC
                TTGGGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAG

                AspProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSer
3481            GATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCC
                CTAGGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGG

                ThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIle
3541            ACCTACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATT
                TGGATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAA

                CysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAsp
3601            TGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGAC
                ACACTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAGCCGTGACAGGAACTG

                GlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySer
3661            CAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCC
                GTTCGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGG

                ValThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIlePro
3721            GTCACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCT
                CAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGA

                PheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCys
3781            TTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGT
```

FIG. 16-5

```
AAAATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACA

                HisSerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAla
3841            CATTCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCC
                GTAAGTTTCTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGG

                ValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValVal
3901            GTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTC
                CACCGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAG

                ValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCys
3961            GTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGC
                CACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACG

                AsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThr
4021            AATACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACA
                TTATGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGT

                IleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGly
4081            ATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGG
                TAGTGCGAGGGGGTCCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCC

                LysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSer
4141            AAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCG
                TTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGC

                SerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGlu
4201            TCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAG
                AGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAGTGCGGGCGGCTC

                ThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHis
4261            ACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCAT
                TGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTA

                LeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSer
4321            CTTGAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCC
                GAACTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGG

                GlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCys
4381            CAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGC
                GTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACG

                AlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeu
4441            GCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTC
                CGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAG

                LysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGlu
4501            AAGCCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAA
                TTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTT

                IleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGlu
4561            ATCACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAG
                TAGTGGGACTGCGTGGGTCAGTGGTTTATGTAGTACTGTACGTACAGCCGGCTGGACCTC

                ValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCys
4621            GTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGC
                CAGCAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAAACCGGCGCATAACG

                LeuSerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIle
4681            CTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATC
                GACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAG

                IleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHis
4741            ATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCAC
                TATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTCGTG
```

FIG. 16-6

```
      LeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGly
4801  TTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGC
      AATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCG

      LeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrp
4861  CTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGG
      GAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACC

      GlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyr
4921  CAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATAC
      GTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATG

      LeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThr
4981  TTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACA
      AACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGT

      AlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGly
5041  GCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGG
      CGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCC

      TrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeu
5101  TGGGTGGCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTA
      ACCCACCGACGGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAAT

      AlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGly
5161  GCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGG
      CGACCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCC

      TyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValPro
5221  TATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCC
      ATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGG

      SerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValVal
5281  TCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTC
      AGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAG

      GlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGln
5341  GGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAG
      CCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTC

      TrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyr
5401  TGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTAC
      ACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGGTGCGTGATG

      ValProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThr
5461  GTGCCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACC
      CACGGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGG

      GlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGly
5521  CAGCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGT
      GTCGAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCA

      SerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrp
5581  TCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGG
      AGGACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACC

      LeuLysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGly
5641  CTAAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGG
      GATTTTCGATTCGAGTACGGTGTCGACGGACCCTAGGGGAAACACAGGACGGTCGCGCCC

      TyrLysGlyValTrpArgValAspGlyIleMetHisThrArgCysHisCysGlyAlaGlu
5701  TATAAGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAG
      ATATTCCCCCAGACCGCTCACCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTC

      IleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsn
5761  ATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAAC
```

FIG. 16-7

```
     TAGTGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTG

     MetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuPro
5821 ATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCT
     TACACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGA

     AlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArg
5881 GCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGG
     CGCGGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATACACCTCTATTCC

     GlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCys
5941 CAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGC
     GTCCACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAGTTTACGGGCACG

     GlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAla
6001 CAGGTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCG
     GTCCAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGC

     ProProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyr
6061 CCCCCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATAC
     GGGGGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATG

     ProValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMet
6121 CCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATG
     GGCCATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTAC

     LeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySer
6181 CTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCA
     GAGTGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGT

     ProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThr
6241 CCCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACT
     GGGGGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGA

     CysThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArg
6301 TGCACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGG
     ACGTGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCC

     GlnGluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAsp
6361 CAGGAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGAC
     GTCCTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTG

     SerPheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIle
6421 TCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATC
     AGGAAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAG

     LeuArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsn
6481 CTGCGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAAC
     GACGCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGGCAAACCCGCGCCGGCCTGATATTG

     ProProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCys
6541 CCCCCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGT
     GGGGGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACA

     ProLeuProProProLysSerProProValProProProArgLysLysArgThrValVal
6601 CCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTC
     GGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAG

     LeuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySer
6661 CTCACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGC
     GAGTGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCG

     SerSerThrSerGlyIleThrGlyAspAsnThrThrThrSerSerGluProAlaProSer
6721 TCCTCAACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCCCCTTCT
     AGGAGTTGAAGGCCGTAATGCCCGCTGTTATGCTGTTGTAGGAGACTCGGGCGGGGAAGA
```

FIG. 16-8

```
                GlyCysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGlu
6781            GGCTGCCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCCTGGAGGGGGAG
                CCGACGGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGGACCTCCCCCTC

                ProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAla
6841            CCTGGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCG
                GGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGC

                GluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCys
6901            GAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGC
                CTCCTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACG

                AlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHis
6961            GCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCAC
                CGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTG

                AsnLeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPhe
7021            AATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT
                TTAAACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAA

                AspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAla
7081            GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCG
                CTGTCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGC

                AlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProPro
7141            GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA
                CGCAGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGT

                HisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLys
7201            CACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG
                GTGAGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTC

                AlaValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIle
7261            GCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATA
                CGGCATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTAT

                AspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArg
7321            GACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT
                CTGTGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCA

                LysProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAla
7381            AAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCT
                TTCGGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGA

                LeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGln
7441            TTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAA
                AACATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTT

                TyrSerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrPro
7501            TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCA
                ATGAGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGT

                MetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArg
7561            ATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGT
                TACCCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCA

                ThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLys
7621            ACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAG
                TGCCTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTC

                SerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCys
7681            TCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGC
                AGGGAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACG

                GlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThr
7741            GGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT
```

FIG. 16-9

```
        CCGATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGA

        CysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeu
7801    TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC
        ACGATGTAGTTCCGGGCCCGTCGGACAGCTCGGCGTCCCGAGGTCCTGACGTGGTACGAG

        ValCysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAla
7861    GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCG
        CACACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGC

        SerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProPro
7921    AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCTGGGGACCCCCCA
        TCGGACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGGGACCCCTGGGGGGT

        GlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHis
7981    CAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC
        GTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTG

        AspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArg
8041    GACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGA
        CTGCCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGTTGGGGGGAGCGCTCT

        AlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMet
8101    GCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG
        CGACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTAC

        PheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIle
8161    TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATA
        AAACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATAT

        AlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIle
8221    GCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATA
        CGGTCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTAT

        GluProLeuAspLeuProProIleIleGlnArgLeu
8281    GAACCACTTGATCTACCTCCAATCATTCAAAGACTC
        CTTGGTGAACTAGATGGAGGTTAGTAAGTTTCTGAG
```

FIG. 17-1

```
-319  CACTCCACCATGAATCACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAAAGCGTCTAG
      GTGAGGTGGTACTTAGTGAGGGGACACTCCTTGATGACAGAAGTGCGTCTTTCGCAGATC

-259  CCATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCCTCCCGGGAGAGCCATA
      GGTACCGCAATCATACTCACAGCACGTCGGAGGTCCTGGGGGGGAGGGCCCTCTCGGTAT

-199  GTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGA
      CACCAGACGCCTTGGCCACTCATGTGGCCTTAACGGTCCTGCTGGCCCAGGAAAGAACCT

-139  TCAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCAAGACTGCTAGCCGAGTAGT
      AGTTGGGCGAGTTACGGACCTCTAAACCCGCACGGGGGCGTTCTGACGATCGGCTCATCA

- 79  GTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAG
      CAACCCAGCGCTTTCCGGAACACCATGACGGACTATCCCACGAACGCTCACGGGGCCCTC

- 19  GTCTCGTAGACCGTGCACC
      CAGAGCATCTGGCACGTGG
                                   Arg   Thr
      MetSerThrAsnProLysProGlnLysLysAsnLysArgAsnThrAsnArgArgProGln
   1  ATGAGCACGAATCCTAAACCTCAAAAAAAAAAACAAACGTAACACCAACCGTCGCCCACAG
      TACTCGTGCTTAGGATTTGGAGTTTTTTTTTTGTTTGCATTGTGGTTGGCAGCGGGTGTC

      AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArg
  61  GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG
      CTGCAGTTCAAGGGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCC

      GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly
 121  GGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGT
      CCGGGATCTAACCCACACGCGCGCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCA

      ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGly
 181  AGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG
      TCTGCAGTCGGATAGGGGTTCCGAGCAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCC

      TyrProTrpProLeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro
 241  TACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC
      ATGGGAACCGGGGAGATACCGTTACTCCCGACGCCCACCCGCCCTACCGAGGACAGAGGG

      ArgGlySerArgProSerTrpGlyProThrAspProArgArgArgSerArgAsnLeuGly
 301  CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT
      GCACCGAGAGCCGGATCGACCCCGGGGTGTCTGGGGGCCGCATCCAGCGCGTTAAACCCA

      LysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuVal
 361  AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC
      TTCCAGTAGCTATGGGAATGCACGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAG

      GlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeuGluAsp
 421  GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC
      CCGCGGGGAGAACCTCCGCGACGGTCCCGGGACCGCGTACCGCAGGCCCAAGACCTTCTG

                                                        Thr
      GlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeuAla
 481  GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC
      CCGCACTTGATACGTTGTCCCTTGGAAGGACCAACGAGAAAGAGATAGAAGGAAGACCGG

      LeuLeuSerCysLeuThrValProAlaSerAlaTyrGlnValArgAsnSerThrGlyLeu
 541  CTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCTT
      GACGAGAGAACGAACTGACACGGGCGAAGCCGGATGGTTCACGCGTTGAGGTGCCCCGAA

      TyrHisValThrAsnAspCysProAsnSerSerIleValTyrGluAlaAlaAspAlaIle
 601  TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC
      ATGGTGCAGTGGTTACTAACGGGATTGAGCTCATAACACATGCTCCGCCGGCTACGGTAG

      LeuHisThrProGlyCysValProCysValArgGluGlyAsnAlaSerArgCysTrpVal
 661  CTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTCGAGGTGTTGGGTG
      GACGTGTGAGGCCCCACGCAGGGAACGCAAGCACTCCCGTTGCGGAGCTCCACAACCCAC
```

FIG. 17-2

```
                AlaMetThrProThrValAlaThrArgAspGlyLysLeuProAlaThrGlnLeuArgArg
 721  GCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGACGCAGCTTCGACGT
      CGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGGCGCTGCGTCGAAGCTGCA

                HisIleAspLeuLeuValGlySerAlaThrLeuCysSerAlaLeuTyrValGlyAspLeu
 781  CACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGACCTA
      GTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACAAGCCGGGAGATGCACCCCCTGGAT

                CysGlySerValPheLeuValGlyGlnLeuPheThrPheSerProArgArgHisTrpThr
 841  TGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCGCCACTGGACG
      ACGCCCAGACAGAAAGAACAGCCGGTTGACAAGTGGAAGAGAGGGTCCGCGGTGACCTGC

                ThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMetAlaTrp
 901  ACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCGCATGGCATGG
      TGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGCCCAGTGGCGTACCGTACC

                                                          Val
                AspMetMetMetAsnTrpSerProThrThrAlaLeuValMetAlaGlnLeuLeuArgIle
 961  GATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAATGGCTCAGCTGCTCCGGATC
      CTATACTACTACTTGACCAGGGGATGCTGCCGCAACCATTACCGAGTCGACGAGGCCTAG

                ProGlnAlaIleLeuAspMetIleAlaGlyAlaHisTrpGlyValLeuAlaGlyIleAla
1021  CCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG
      GGTGTTCGGTAGAACCTGTACTAGCGACCACGAGTGACCCCTCAGGACCGCCCGTATCGC

                TyrPheSerMetValGlyAsnTrpAlaLysValLeuValValLeuLeuLeuPheAlaGly
1081  TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC
      ATAAAGAGGTACCACCCCTTGACCCGCTTCCAGGACCATCACGACGACGATAAACGGCCG

                ValAspAlaGluThrHisValThrGlyGlySerAlaGlyHisThrValSerGlyPheVal
1141  GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCACACTGTGTCTGGATTTGTT
      CAGCTGCGCCTTTGGGTGCAGTGGCCCCCTTCACGGCCGGTGTGACACAGACCTAAACAA

                SerLeuLeuAlaProGlyAlaLysGlnAsnValGlnLeuIleAsnThrAsnGlySerTrp
1201  AGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGATCAACACCAACGGCAGTTGG
      TCGGAGGAGCGTGGTCCGCGGTTCGTCTTGCAGGTCGACTAGTTGTGGTTGCCGTCAACC

                HisLeuAsnSerThrAlaLeuAsnCysAsnAspSerLeuAsnThrGlyTrpLeuAlaGly
1261  CACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAACACCGGCTGGTTGGCAGGG
      GTGGAGTTATCGTGCCGGGACTTGACGTTACTATCGGAGTTGTGGCCGACCAACCGTCCC

                LeuPheTyrHisHisLysPheAsnSerSerGlyCysProGluArgLeuAlaSerCysArg
1321  CTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGA
      GAAAAGATAGTGGTGTTCAAGTTGAGAAGTCCGACAGGACTCTCCGATCGGTCGACGGCT

                ProLeuThrAspPheAspGlnGlyTrpGlyProIleSerTyrAlaAsnGlySerGlyPro
1381  CCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTATGCCAACGGAAGCGGCCCC
      GGGGAATGGCTAAAACTGGTCCCGACCCCGGGATAGTCAATACGGTTGCCTTCGCCGGGG

                AspGlnArgProTyrCysTrpHisTyrProProLysProCysGlyIleValProAlaLys
1441  GACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAACCTTGCGGTATTGTGCCCGCGAAG
      CTGGTCGCGGGGATGACGACCGTGATGGGGGGTTTTGGAACGCCATAACACGGGCGCTTC

                SerValCysGlyProValTyrCysPheThrProSerProValValValGlyThrThrAsp
1501  AGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACGACCGAC
      TCACACACACCAGGCCATATAACGAAGTGAGGGTCGGGGCACCACCACCCTTGCTGGCTG

                ArgSerGlyAlaProThrTyrSerTrpGlyGluAsnAspThrAspValPheValLeuAsn
1561  AGGTCGGGCGCGCCCACCTACAGCTGGGGTGAAAATGATACGGACGTCTTCGTCCTTAAC
      TCCAGCCCGCGCGGGTGGATGTCGACCCCACTTTTACTATGCCTGCAGAAGCAGGAATTG

                AsnThrArgProProLeuGlyAsnTrpPheGlyCysThrTrpMetAsnSerThrGlyPhe
1621  AATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACTGGATTC
      TTATGGTCCGGTGGCGACCCGTTAACCAAGCCAACATGGACCTACTTGAGTTGACCTAAG
```

FIG. 17-3

```
                ThrLysValCysGlyAlaProProCysValIleGlyGlyAlaGlyAsnAsnThrLeuHis
1681            ACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGGGCGGGCAACAACACCCTGCAC
                TGGTTTCACACGCCTCGCGGAGGAACACAGTAGCCTCCCCGCCCGTTGTTGTGGGACGTG

                CysProThrAspCysPheArgLysHisProAspAlaThrTyrSerArgCysGlySerGly
1741            TGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATACTCTCGGTGCGGCTCCGGT
                ACGGGGTGACTAACGAAGGCGTTCGTAGGCCTGCGGTGTATGAGAGCCACGCCGAGGCCA

                      Leu
                ProTrpIleThrProArgCysLeuValAspTyrProTyrArgLeuTrpHisTyrProCys
1801            CCCTGGATCACACCCAGGTGCCTGGTCGACTACCCGTATAGGCTTTGGCATTATCCTTGT
                GGGACCTAGTGTGGGTCCACGGACCAGCTGATGGGCATATCCGAAACCGTAATAGGAACA

                ThrIleAsnTyrThrIlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeu
1861            ACCATCAACTACACCATATTTAAAATCAGGATGTACGTGGGAGGGGTCGAACACAGGCTG
                TGGTAGTTGATGTGGTATAAATTTTAGTCCTACATGCACCCTCCCCAGCTTGTGTCCGAC

                GluAlaAlaCysAsnTrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSer
1921            GAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCC
                CTTCGACGGACGTTGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGG

                GluLeuSerProLeuLeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThr
1981            GAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACA
                CTCGAGTCGGGCAATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGT

                ThrLeuProAlaLeuSerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGln
2041            ACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTGGACGTGCAG
                TGGGATGGTCGGAACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTC

                TyrLeuTyrGlyValGlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValVal
2101            TACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTT
                ATGAACATGCCCCACCCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCTCATGCAGCAA

                LeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeu
2161            CTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTC
                GAGGACAAGGAAGACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAG

                IleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAla
2221            ATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCC
                TATAGGGTTCGCCTCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGG

                GlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrpTyrLeuLysGly
2281            GGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTTGCATGGTATTTGAAGGGT
                CCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAACGTACCATAAACTTCCCA

                LysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeuLeuLeuLeuLeu
2341            AAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGGCCTCTCCTCCTGCTCCTG
                TTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCCTACACCGGAGAGGAGGACGAGGAC

                LeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAlaSerCysGlyGly
2401            TTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCGTCGTGTGGCGGT
                AACCGCAACGGGGTCGCCCGCATGCGCGACCTGTGCCTCCACCGGCGCAGCACACCGCCA

                ValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyrIleSer
2461            GTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACAAGCGCTATATCAGC
                CAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATAATGTTCGCGATATAGTCG

                                  Asn
                TrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHisValTrp
2521            TGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCAACTGCACGTGTGG
                ACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTTCGCGTTGACGTGCACACC

                IleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeuMetCysAlaVal
2581            ATTCCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTCATGTGTGCTGTA
                TAAGGGGGGGAGTTGCAGGCTCCCCCCGCGCTGCGGCAGTAGAATGAGTACACACGACAT
```

FIG. 17-4

```
                HisProThrLeuValPheAspIleThrLysLeuLeuLeuAlaValPheGlyProLeuTrp
2641            CACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGTCTTCGGACCCCTTTGG
                GTGGGCTGAGACCATAAACTGTAGTGGTTTAACGACGACCGGCAGAAGCCTGGGGAAACC

                IleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGlnGlyLeuLeuArg
2701            ATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAAGGCCTTCTCCGG
                TAAGAAGTTCGGTCAAACGAATTTCATGGGATGAAACACGCGCAGGTTCCGGAAGAGGCC

                PheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMetValIleIleLys
2761            TTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATCATTAAG
                AAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCACGTTTACCAGTAGTAATTC

                LeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAla
2821            TTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGACTGGGCG
                AATCCCCGCGAATGACCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTGACCCGC

                HisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPheSerGlnMetGlu
2881            CACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAAATGGAG
                GTGTTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTTTACCTC

                ThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeu
2941            ACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATCATCAACGGCTTG
                TGGTTCGAGTAGTGCACCCCCCGTCTATGGCGGCGCACGCCACTGTAGTAGTTGCCGAAC

                ProValSerAlaArgArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSer
3001            CCTGTTTCCGCCCGCAGGGGCCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCC
                GGACAAAGGCGGGCGTCCCCGGCCCTCTATGACGAGCCCGGTCGGCTACCTTACCAGAGG

                LysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeu
3061            AAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTA
                TTCCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCGGGTCGTCTGTTCCCCGGAGGAT

                GlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGln
3121            GGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAG
                CCCACGTATTAGTGGTCGGATTGACCGGCCCTGTTTTTGGTTCACCTCCCACTCCAGGTC

                IleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThr
3181            ATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACT
                TAACACAGTTGACGACGGGTTTGGAAGGACCGTTGCACGTAGTTACCCCACACGACCTGA

                ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMet
3241            GTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATG
                CAGATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTCCCAGGACAGTAGGTCTAC

                                                    Ser       Thr
                TyrThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeu
3301            TATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTG
                ATATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAAC

                ThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIle
3361            ACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATT
                TGTGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGCGGCTACAGTAA

                ProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyr
3421            CCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTAC
                GGGCACGCGGCCGCCCCACTATCGTCCCCGTCGGACGACAGCGGGGCCGGGTAAAGGATG

                LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe
3481            TTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTT
                AACTTTCCGAGGAGCCCCCCAGGCGACAACACGGGGCGCCCCGTGCGGCACCCGTATAAA

                ArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIleProValGluAsn
3541            AGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAAC
                TCCCGGCGCCACACGTGGGCACCTCACCGATTCCGCCACCTGAAATAGGGACACCTCTTG
```

FIG. 17-5

```
                LeuGluThrThrMetArgSerProValPheThrAspAsnSerSerProProValValPro
3601            CTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCACCAGTAGTGCCC
                GATCTCTGTTGGTACTCCAGGGGCCACAAGTGCCTATTGAGGAGAGGTGGTCATCACGGG

                GlnSerPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysVal
3661            CAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAAAGCACCAAGGTC
                GTCTCGAAGGTCCACCGAGTGGAGGTACGAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAG

                ProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAla
3721            CCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCCTCTGTTGCTGCA
                GGCCGACGTATACGTCGAGTCCCGATATTCCACGATCATGAGTTGGGGAGACAACGACGT

                                                                Leu
                ThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThr
3781            ACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACC
                TGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCCTGG

                GlyValArgThrIleThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeu
3841            GGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTT
                CCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAA

                AlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSer
3901            GCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCC
                CGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACACTGCTCACGGTGAGG

                ThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGly
3961            ACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCAGAGACTGCGGGG
                TGCCTACGGTGTAGGTAGAACCCGTAGCCGTGACAGGAACTGGTTCGTCTCTGACGCCCC

                AlaArgLeuValValLeuAlaThrAlaThrProProGlySerValThrValProHisPro
4021            GCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTCACTGTGCCCCATCCC
                CGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGTGACACGGGGTAGGG

                AsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIle
4081            AACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTTTACGGCAAGGCTATC
                TTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAAATGCCGTTCCGATAG

                ProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCys
4141            CCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCATTCAAAGAAGAAGTGC
                GGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTAAGTTTCTTCTTCACG

                AspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGly
4201            GACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGT
                CTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCA

                LeuAspValSerValIleProThrSerGlyAspValValValValAlaThrAspAlaLeu
4261            CTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCAACCGATGCCCTC
                GAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCACCGTTGGCTACGGGAG

                                                        Tyr
                MetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGln
4321            ATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAATACGTGTGTCACCCAG
                TACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTATGCACACAGTGGGTC

                                                    Ser
                ThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuProGlnAsp
4381            ACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGAT
                TGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAGTGCGAGGGGGTCCTA

                AlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArg
4441            GCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCAGGCATCTACAGA
                CGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCCTTCGGTCCGTAGATGTCT

                PheValAlaProGlyGluArgProSerGlyMetPheAspSerSerValLeuCysGluCys
4501            TTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTCCTCTGTGAGTGC
                AAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGCAGGCAGGAGACACTCACG
```

FIG. 17-6

```
                TyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArg
4561            TATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACAGTTAGGCTACGA
                ATACTGCGTCCGACACGAACCATACTCGAGTGCGGGCGGCTCTGATGTCAATCCGATGCT

                AlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGly
4621            GCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAATTTTGGGAGGGC
                CGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAACTTAAAACCCTCCCG

                ValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSerGly
4681            GTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGG
                CAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCC

                GluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaPro
4741            GAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCT
                CTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGATCCCGAGTTCGGGGA

                ProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGly
4801            CCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCCACCCTCCATGGG
                GGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTCGGGTGGGAGGTACCC

                ProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisPro
4861            CCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCACCCTGACGCACCCA
                GGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGT

                ValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrp
4921            GTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTCACGAGCACCTGG
                CAGTGGTTTATGTAGTACTGTACGTACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACC

                ValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysVal
4981            GTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCAACAGGCTGCGTG
                CACGAGCAACCGCCGCAGGACCGACGAAACCGGCGCATAACGGACAGTTGTCCGACGCAC

                ValIleValGlyArgValValLeuSerGlyLysProAlaIleIleProAspArgGluVal
5041            GTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTC
                CAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAG

                LeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGln
5101            CTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAA
                GAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTT

                GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSer
5161            GGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTGCAGACCGCGTCC
                CCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGG

                ArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPhe
5221            CGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTTC
                GCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAG

                TrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThr
5281            TGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGGCGGGCTTGTCAACG
                ACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAACCGCCCGAACAGTTGC

                LeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerPro
5341            CTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCTGTCACCAGCCCA
                GACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGACGACAGTGGTCGGGT

                LeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeu
5401            CTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTGGCTGCCCAGCTC
                GATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACCCACCGACGGGTCGAG

                AlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGly
5461            GCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGCGCCGCCATCGGC
                CGGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAATCGACCGCGGCGGTAGCCG

                SerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAla
```

FIG. 17-7

```
5521  AGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATGGCGCGGGCGTGGCG
      TCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATACCGCGCCCGCACCGC

                                                              Gly
      GlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSerThrGluAspLeuVal
5581  GGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACGGAGGACCTGGTC
      CCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGGAGGTGCCTCCTGGACCAG

      AsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGlyValValCysAlaAla
5641  AATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCA
      TTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCGCACCAGACACGTCGT

      IleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIle
5701  ATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATGAACCGGCTGATA
      TATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACCTACTTGGCCGACTAT

      AlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrValProGluSerAspAla
5761  GCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCGGAGAGCGATGCA
      CGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCACGGCCTCTCGCTACGT

                                  HisCys
      AlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeu
5821  GCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTG
      CGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTCGAGGACTCCGCTGAC

      HisGlnTrpIleSerSerGluCysThrThrProCysSerGlySerTrpLeuArgAspIle
5881  CACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGGCTAAGGGACATC
      GTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAG

      TrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMet
5941  TGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAAGCTAAGCTCATG
      ACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGATTTTCGATTCGAGTAC

      ProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArg
6001  CCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAGGGGGTCTGGCGA
      GGTGTCGACGGACCCTAGGGGAAACACAGGACGGTCGCGCCCATATTCCCCCAGACCGCT

      Gly
      ValAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLys
6061  GTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAA
      CACCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTT

      AsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPhe
6121  AACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTC
      TTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTACACCTCACCCTGGAAG

      ProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPhe
6181  CCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCGAACTACACGTTC
      GGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGCGGCTTGATGTGCAAG

      AlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHis
6241  GCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTGGGGGACTTCCAC
      CGCGATACCTCCCACAGACGTCTCCTTATACACCTCTATTCCGTCCACCCCCTGAAGGTG

      TyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGlnValProSerProGlu
6301  TACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGGTCCCATCGCCCGAA
      ATGCACTGCCCATACTGATGACTGTTAGAGTTTACGGGCACGGTCCAGGGTAGCGGGCTT

      PhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaProProCysLysProLeu
6361  TTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCCTGCAAGCCCTTG
      AAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGGGGGGACGTTCGGGAAC

      LeuArgGluGluValSerPheArgValGlyLeuHisGluTyrProValGlySerGlnLeu
6421  CTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGGTAGGGTCGCAATTA
      GACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGCCATCCCAGCGTTAAT
```

FIG. 17-8

```
                ProCysGluProGluProAspValAlaValLeuThrSerMetLeuThrAspProSerHis
6481            CCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCACTGATCCCTCCCAT
                GGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTACGAGTGACTAGGGAGGGTA

                IleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerProProSerValAlaSer
6541            ATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCCTCTGTGGCCAGC
                TATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGGGGGAGACACCGGTCG

                SerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHisAsp
6601            TCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACCGCTAACCATGAC
                AGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACGTGGCGATTGGTACTG

                SerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsn
6661            TCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAGATGGGCGGCAAC
                AGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTCCTCTACCCGCCGTTG

                IleThrArgValGluSerGluAsnLysValValIleLeuAspSerPheAspProLeuVal
6721            ATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTCGATCCGCTTGTG
                TAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGGAAGCTAGGCGAACAC

                AlaGluGluAspGluArgGluIleSerValProAlaGluIleLeuArgLysSerArgArg
6781            GCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGGAAGTCTCGGAGA
                CGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGACGCCTTCAGAGCCTCT

                PheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnProProLeuValGluThr
6841            TTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCGCTAGTGGAGACG
                AAGCGGGTCCGGGACGGGCAAACCCGCGCCGGCCTGATATTGGGGGGCGATCACCTCTGC

                TrpLysLysProAspTyrGluProProValValHisGlyCysProLeuProProProLys
6901            TGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTTCCACCTCCAAAG
                ACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACAGGCGAAGGTGGAGGTTTC

                SerProProValProProProArgLysLysArgThrValValLeuThrGluSerThrLeu
6961            TCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACTGAATCAACCCTA
                AGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAGTGACTTAGTTGGGAT

                                    Ser
                SerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSerSerThrSerGlyIle
7021            TCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCAACTTCCGGCATT
                AGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGGAGTTGAAGGCCGTAA

                ThrGlyAspAsnThrThrThrSerSerGluProAlaProSerGlyCysProProAspSer
7081            ACGGGCGACAATACGACAACATCCTCTGAGCCCGCCCCTTCTGGCTGCCCCCCCGACTCC
                TGCCCGCTGTTATGCTGTTGTAGGAGACTCGGGCGGGGAAGACCGACGGGGGGGCTGAGG

                              PheAla
                AspAlaGluSerTyrSerSerMetProProLeuGluGlyGluProGlyAspProAspLeu
7141            GACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAGGGGGAGCCTGGGGATCCGGATCTT
                CTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAA

                SerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGluAspValValCysCys
7201            AGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAGGATGTCGTGTGCTGC
                TCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTCCTACAGCACACGACG

                SerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAlaAlaGluGluGlnLys
7261            TCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCCGCGGAAGAACAGAAA
                AGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGGCGCCTTCTTGTCTTT

                LeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsnLeuValTyrSerThr
7321            CTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAATTTGGTGTATTCCACC
                GACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTGTTAAACCACATAAGGTGG

                ThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAspArgLeuGlnValLeu
7381            ACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGACAGACTGCAAGTTCTG
                TGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTGTCTGACGTTCAAGAC
```

FIG. 17-9

```
                AspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAlaSerLysValLysAla
7441            GACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCGTCAAAAGTGAAGGCT
                CTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGCAGTTTTCACTTCCGA

                   Phe
                AsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHisSerAlaLysSerLys
7501            AACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCACACTCAGCCAAATCCAAG
                TTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGTGTGAGTCGGTTTAGGTTC

                PheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAlaValThrHisIleAsn
7561            TTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCCGTAACCCACATCAAC
                AAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTCCGGCATTGGGTGTAGTTG

                SerValTrpLysAspLeuLeuGluAspAsnValThrProIleAspThrThrIleMetAla
7621            TCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATAGACACTACCATCATGGCT
                AGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTGTGATGGTAGTACCGA

                LysAsnGluValPheCysValGlnProGluLysGlyGlyArgLysProAlaArgLeuIle
7681            AAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAGCCAGCTCGTCTCATC
                TTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCATTCGGTCGAGCAGAGTAG

                ValPheProAspLeuGlyValArgValCysGluLysMetAlaLeuTyrAspValValThr
7741            GTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCTTTGTACGACGTGGTTACA
                CACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAACATGCTGCACCAATGT

                LysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArg
7801            AAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGG
                TTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTTATGAGTGGTCCTGTCGCC

                ValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAsp
7861            GTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATGGGGTTCTCGTATGAT
                CAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTACCCCAAGAGCATACTA

                ThrArgCysPheAspSerThrValThrGluSerAspIleArgThrGluGluAlaIleTyr
7921            ACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTAC
                TGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGCCTCCTCCGTTAGATG

                GlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeu
7981            CAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCCCTCACCGAGAGGCTT
                GTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGGGAGTGGCTCTCCGAA

                                   Gly
                TyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArg
8041            TATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGCTATCGCAGGTGCCGC
                ATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCGATAGCGTCCACGGCG

                AlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArg
8101            GCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGCTACATCAAGGCCCGG
                CGCTCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACGATGTAGTTCCGGGCC

                AlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeu
8161            GCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTGTGTGGCGACGACTTA
                CGTCGGACAGCTCGGCGTCCCGAGGTCCTGACGTGGTACGAGCACACACCGCTGCTGAAT

                ValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSerLeuArgAlaPheThr
8221            GTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCGAGCCTGAGAGCCTTCACG
                CAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCGGACTCTCGGAAGTGC

                GluAlaMetThrArgTyrSerAlaProProGlyAspProProGlnProGluTyrAspLeu
8281            GAGGCTATGACCAGGTACTCCGCCCCCCCTGGGGACCCCCCACAACCAGAATACGACTTG
                CTCCGATACTGGTCCATGAGGCGGGGGGGACCCCTGGGGGGTGTTGGTCTTATGCTGAAC

                GluLeuIleThrSerCysSerSerAsnValSerValAlaHisAspGlyAlaGlyLysArg
8341            GAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGACGGCGCTGGAAAGAGG
                CTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTGCCGCGACCTTTCTCC
```

```
                ValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAlaAlaTrpGluThrAla
8401            GTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGAGCTGCGTGGGAGACAGCA
                CAGATGATGGAGTGGGCACTGGGATGTTGGGGGGAGCGCTCTCGACGCACCCTCTGTCGT

                ArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPheAlaProThrLeuTrp
8461            AGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTTGCCCCCACACTGTGG
                TCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAACGGGGGTGTGACACC

                AlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAlaArgAspGlnLeuGlu
8521            GCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCCAGGGACCAGCTTGAA
                CGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGGTCCCTGGTCGAACTT

                GlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuPro
8581            CAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAACCACTTGATCTACCT
                GTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTTGGTGAACTAGATGGA

                ProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHisSerTyrSerProGly
8641            CCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCACAGTTACTCTCCAGGT
                GGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTGTCAATGAGAGGTCCA

                GluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValProProLeuArgAlaTrp
8701            GAAATTAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTACCGCCCTTGCGAGCTTGG
                CTTTAATTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGCGGGAACGCTCGAACC

                Gly
                ArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGlyGlyArgAlaAlaIle
8761            AGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGAGGAGGCAGGGCTGCCATA
                TCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCTCCGTCCCGACGGTAT

                CysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
8821            TGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAAC
                ACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTCGAGTTTG
```

FIG. 17-10

FIG. 18

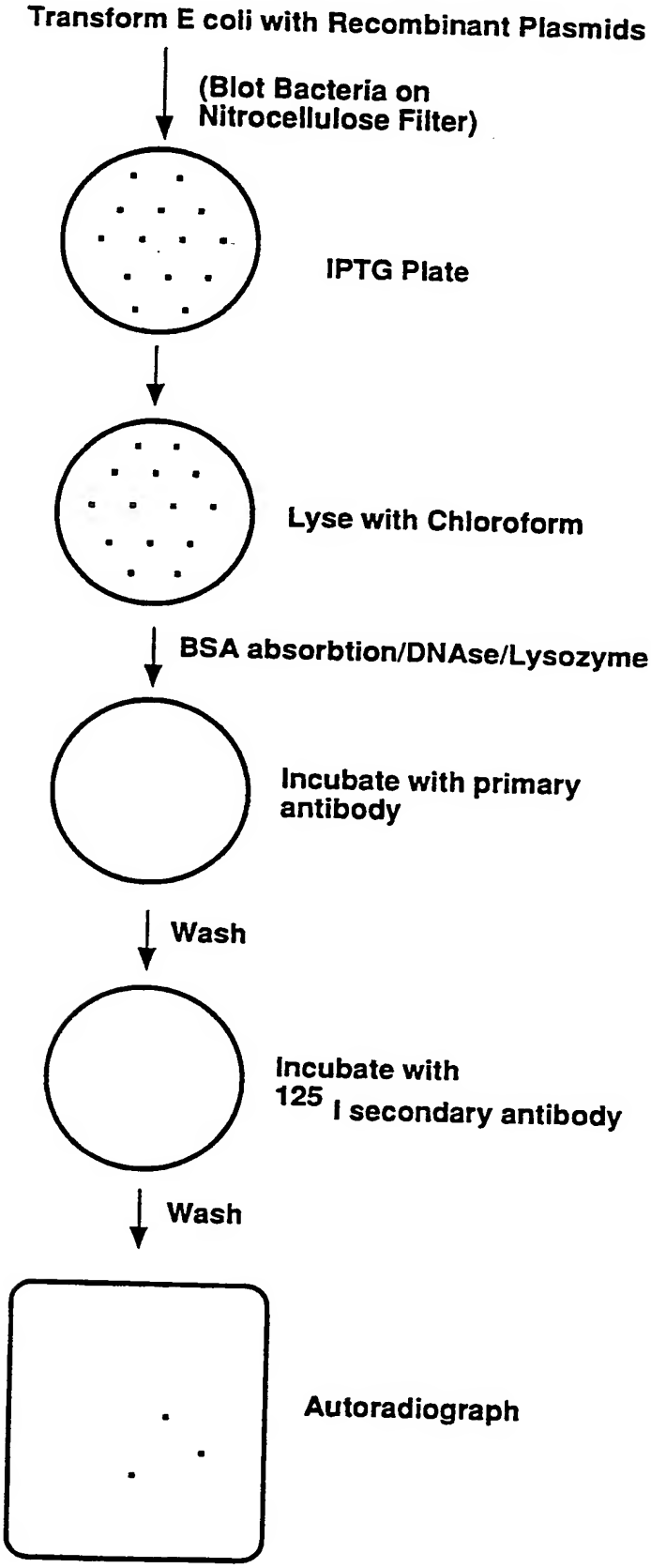

IMMUNOLOGICAL SCREENING IN BACTERIA
Transform E coli with Recombinant Plasmids
(Blot Bacteria on Nitrocellulose Filter)
IPTG Plate
Lyse with Chloroform
BSA absorbtion/DNAse/Lysozyme
Incubate with primary antibody
Wash
Incubate with 125 I secondary antibody
Wash
Autoradiograph

FIG. 19-1

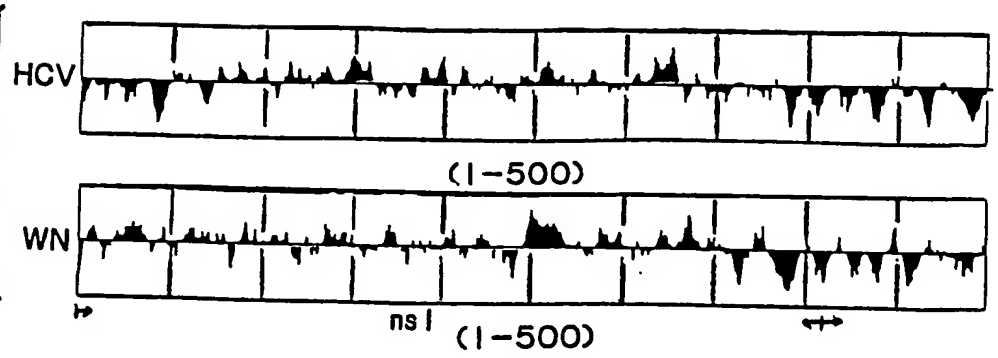


FIG. 19-2

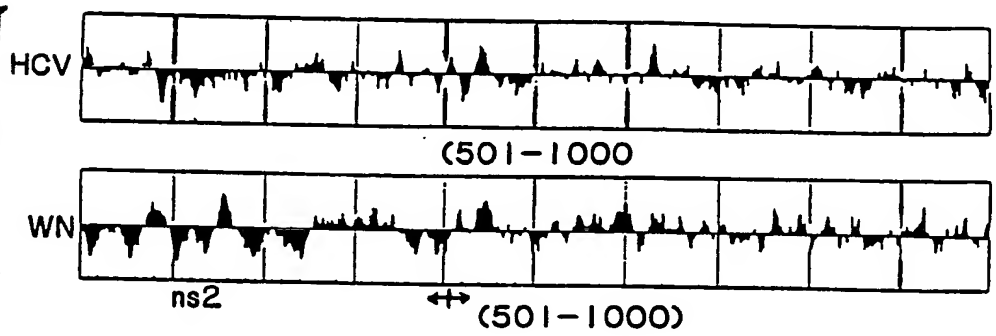


FIG. 19-3

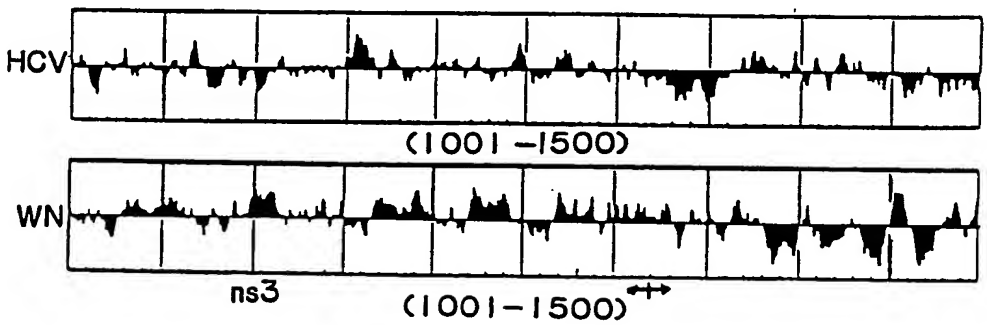


FIG. 19-4

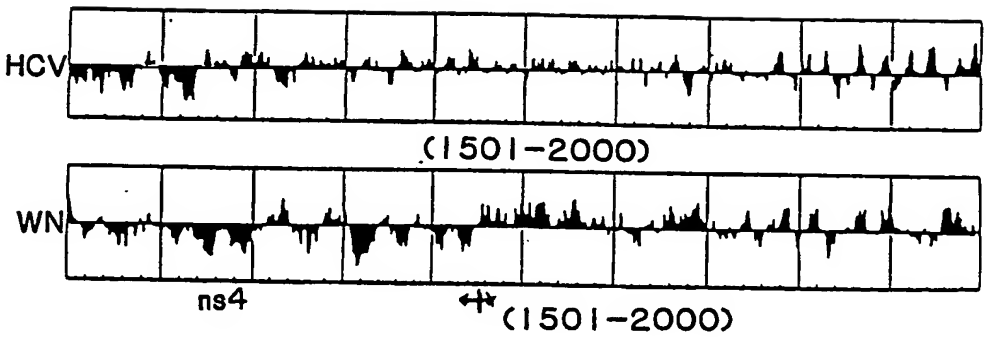
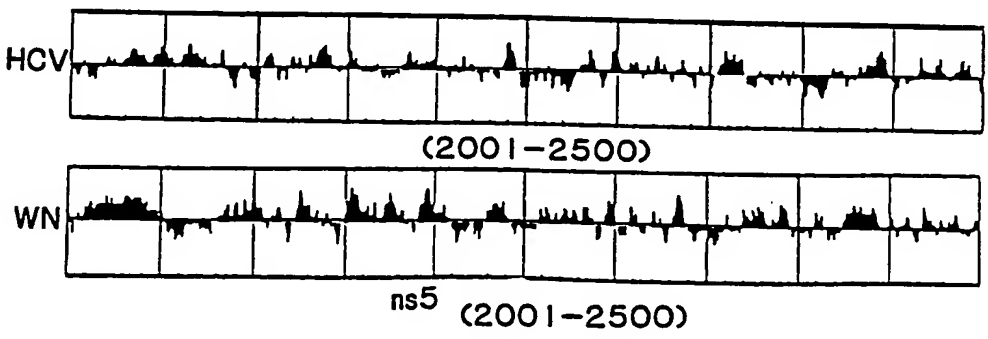


FIG. 19-5



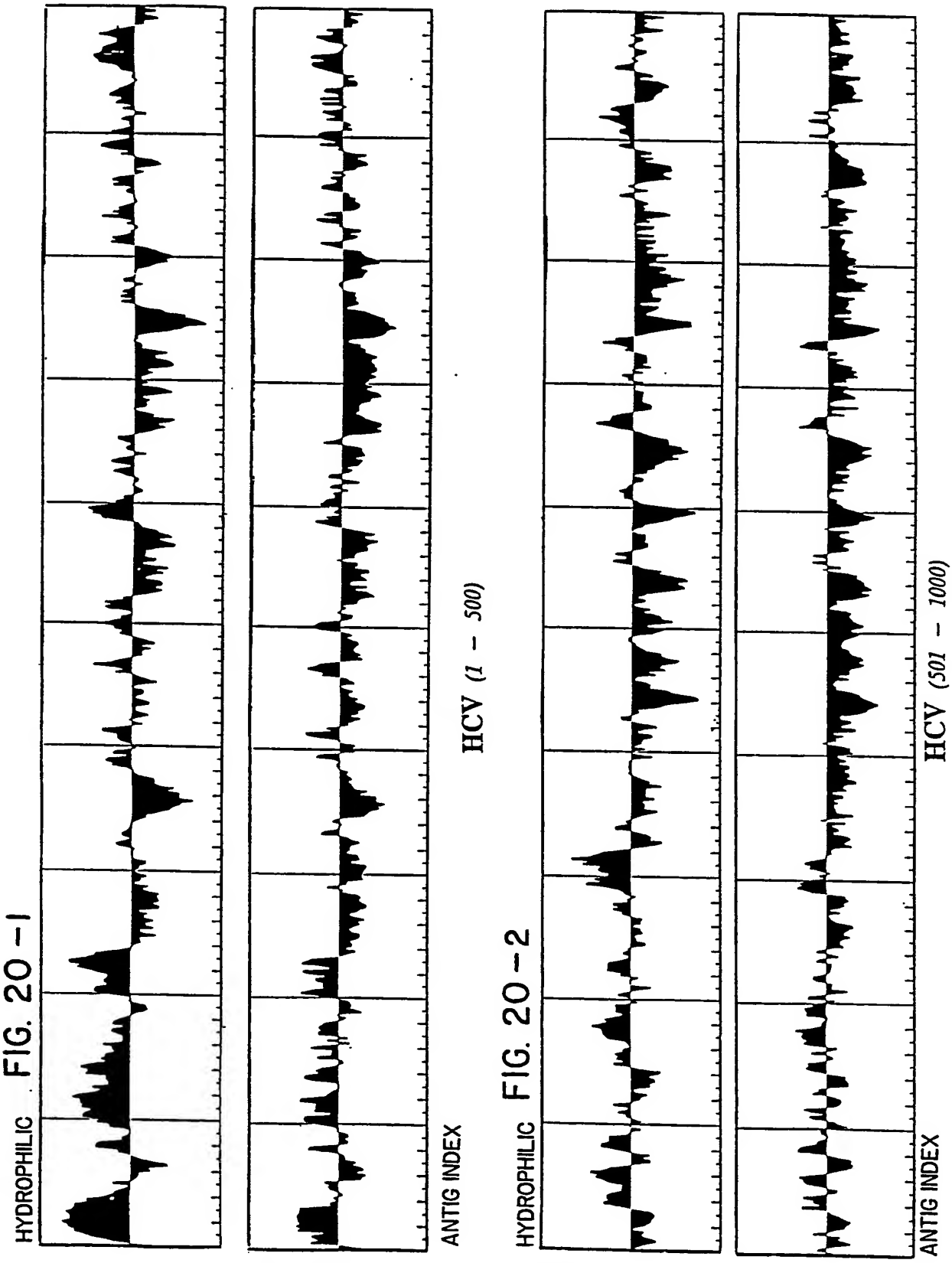


FIG. 20-1

FIG. 20-2

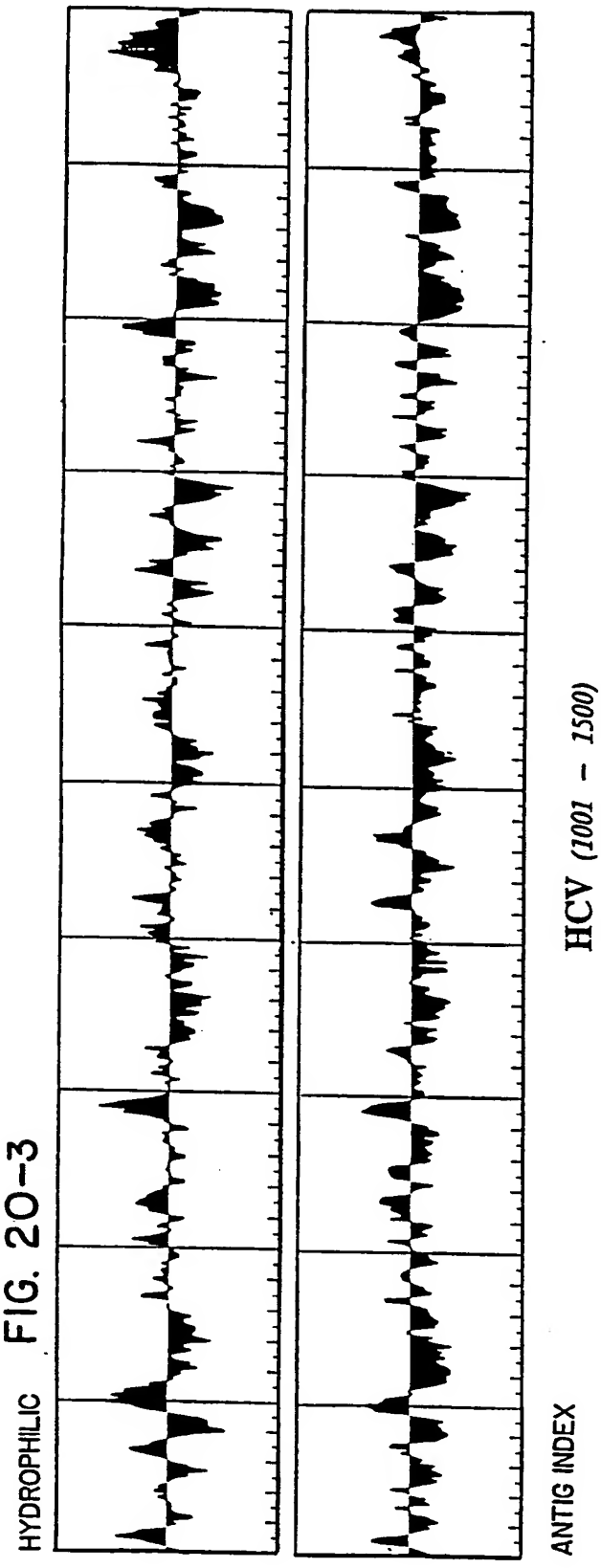


FIG. 20-3

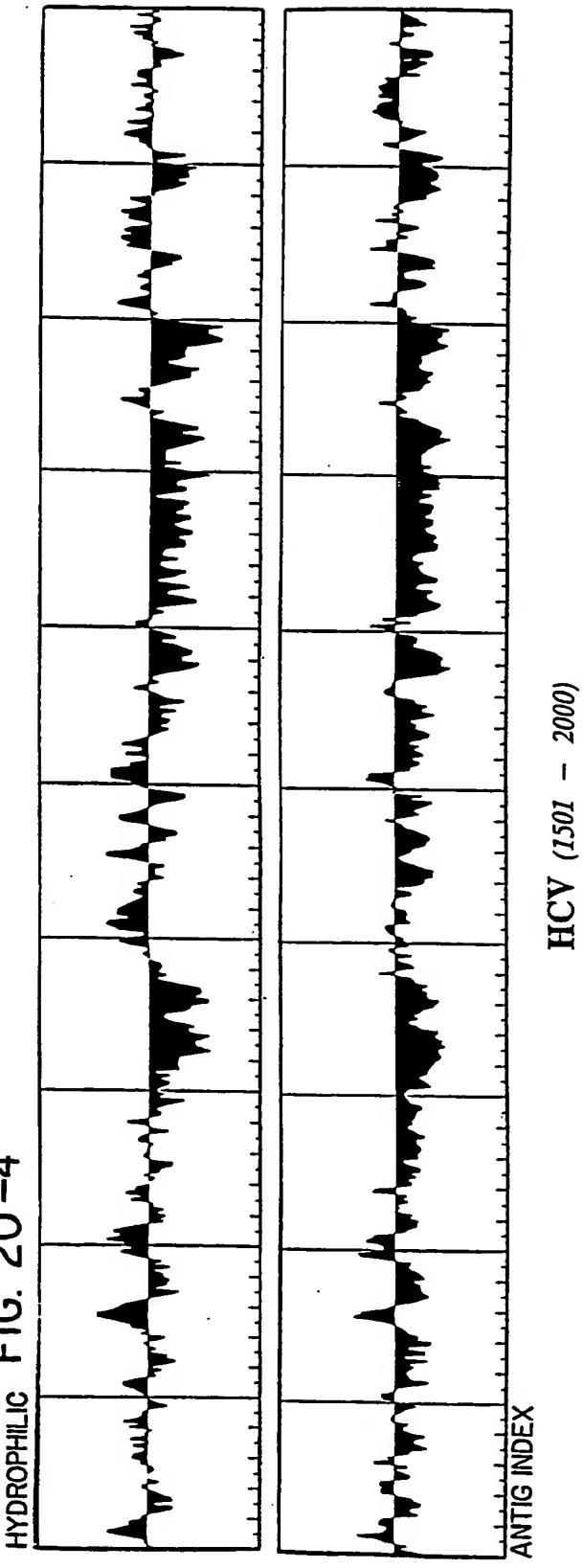


FIG. 20-4

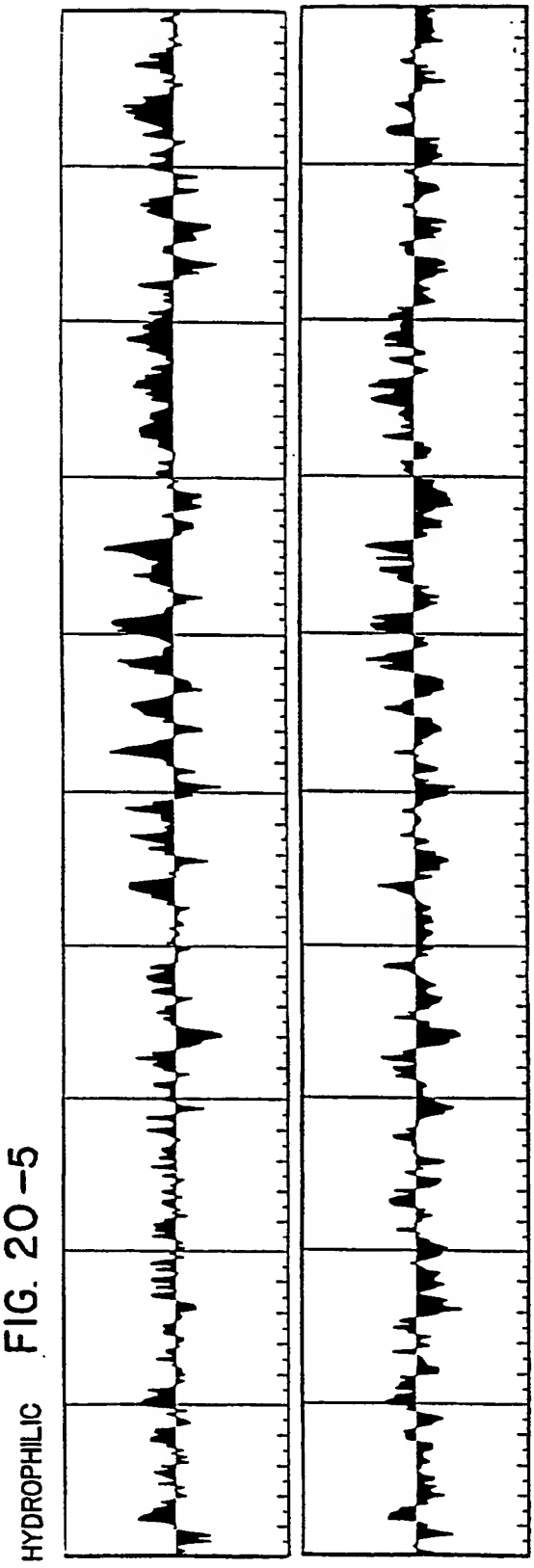


FIG. 20-5

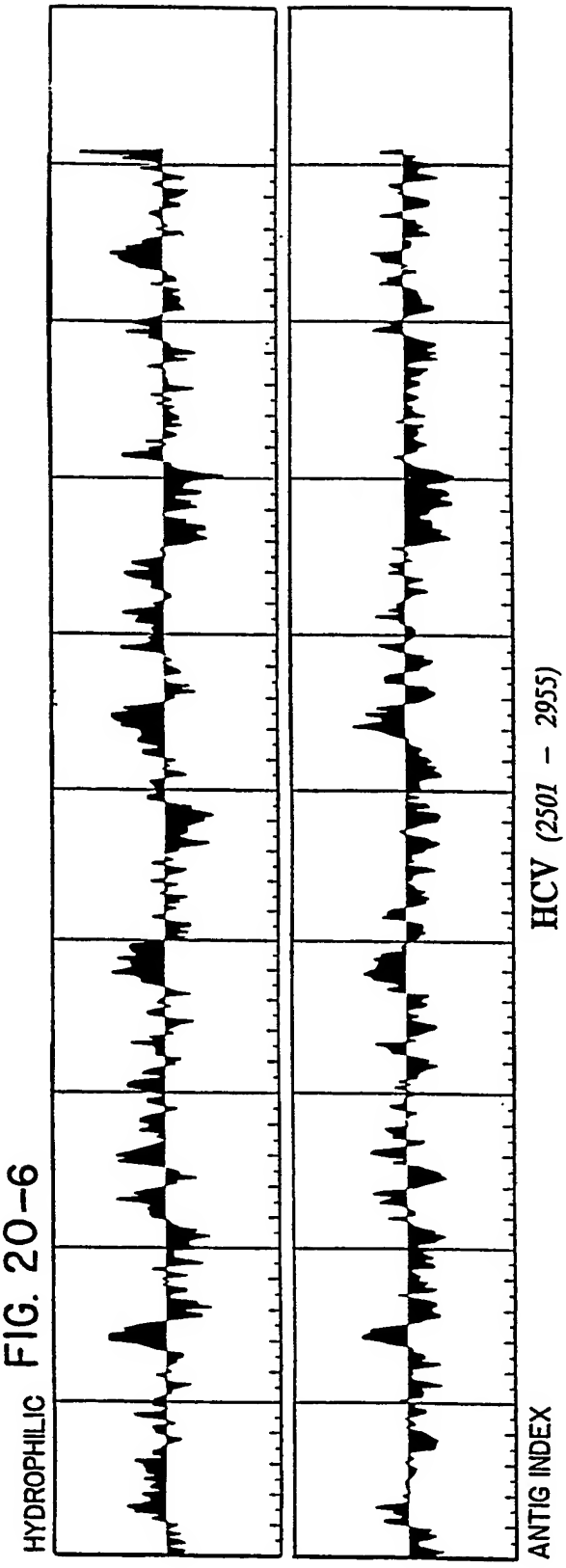


FIG. 20-6

Some conserved co-linear peptides in HCV & Flaviviruses

	NS3 region	NS5 Highly-conserved Polymerase region
Flaviviruses (Yellow Fever, West Nile, Dengue)	TATPPG---------SAAQRRGRIGRNP------	----------GDDCVV
	****** * ****** **	*** **
HCV	TATPPG---------SRTQRRGRTGRGK------	----------GDDLVV
	#1348 #1483	#2737

FIG. 21

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US90/01348

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 6

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5):A61K 39/12;C07H 21/04;C07K 7/06,08,10,13/00,15/28;
US: 536/27;435/6,7,69.1,320,240.1;530/324-327,350,387,416;424/89

II. FIELDS SEARCHED

see attachment

Minimum Documentation Searched 7

Classification System	Classification Symbols
U.S.	536/27;435/6,7,69.1,320,240.1 530/324-327,350,387,416; 424/89

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched 8

Automated Patent Search, Chemical Abstract Service

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
A,P	US, A, 4,870,026 WANDS ET AL. 26 September 1989	1-34
A	US, A, 4,702,909 VILLAREJOS ET AL. 27 October 1987	1-34
A	US, A, 4,542,016 TREPO 17 September 1985	1-34
X,P	Science, "Isolation of a c DNA clone derived from a blood-borne non-A, non-B viral hepatitis genone" volume 244, pp 359-362. CHOO ET AL. 21 April 1989 see abstract & figs.	1-12,15,16, 18-20,22,28-31,33
X,P	Science,"An assay for circulating antibodies to a major ehologic virus of human nonA,nonB Hepatitis" volume 244, pp 362-364. KUO ET AL 21 April 1989 see abstract.	18,19,22

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
28 MAY 1990	13 JUL 1990
International Searching Authority	**Signature of Authorized Officer**
ISA/US	NINA OSSANNA

Form PCT/ISA/210 (second sheet) (Rev.11-87)

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Continuation fo Classification:

IPC(5) C12N 1/11,15/02,51 C12P 21/02; C12Q 1/68; G01N 33/53

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers . . , because they relate to subject matter [12] not required to be searched by this Authority, namely:

2. ☐ Claim numbers..... . .. , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
A61K 37/00, 37/02, 37/20
A61K 47/00, 47/06, C08H 61/00
C08H 61/12, C08L 65/02
C12N 5/00, 5/06, 5/08
C12N 5/16, 5/22, 15/00
C12N 15/06, 15/07, 15/11

A1

(11) International Publication Number: WO 93/19768

(43) International Publication Date: 14 October 1993 (14.10.93)

(21) International Application Number: PCT/US93/03406

(22) International Filing Date: 5 April 1993 (05.04.93)

(30) Priority data:
07/864,876 3 April 1992 (03.04.92) US
07/913,669 14 July 1992 (14.07.92) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 300 Lakeside Drive, 22nd Floor, Oakland, CA 94612-3550 (US).

(72) Inventors: SZOKA, Francis, C., Jr. ; 45 Mendosa Avenue, San Francisco, CA 94116 (US). HAENSLER, Jean ; 1803 Judah Street, #2, San Francisco, CA 94112 (US).

(74) Agents: FISHER, Stanley, P. et al.; Fisher & Amzel, 1320 Harbor Bay Parkway, Suite 225, Alameda, CA 94501 (US).

(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SELF-ASSEMBLING POLYNUCLEOTIDE DELIVERY SYSTEM

(57) Abstract

This invention provides a self-assembling polynucleotide delivery system comprising components aiding in the delivery of the polynucleotide to the desired address which are associated via noncovalent interactions with the polynucleotide. The components of this system include DNA-masking components, cell recognition components, charge-neutralization and membrane-permeabilization components, and subcellular localization components. Specific compounds useful in this system are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

SELF-ASSEMBLING POLYNUCLEOTIDE DELIVERY SYSTEM

Technical Field

This invention is in the field of oligonucleotide delivery and gene therapy. In particular this invention is directed to a self-assembling polynucleotide delivery system comprising components aiding in the delivery of the polynucleotide to the desired address which are associated via noncovalent interactions with the polynucleotide. The components of this system include DNA-masking components, cell recognition components, charge-neutralization and membrane-permeabilization components, and subcellular localization components.

Background Art

Cystic Fibrosis (CF) is a fatal recessive genetic disease characterized by abnormalities in chloride transport (McPherson & Dorner, 1991). The locus of the disease has been traced to mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR). J.R. Riordan et al., Science (1989) 245:1066-1073; B. Kerem et al., Science (1989) 245:1073-1080. Correction of the underlying gene defect by complementation or replacement of the defective CFTR is the ultimate cure for CF. Gene therapy, the in vivo delivery and expression of genes, is a fast-developing science that can be used to replace defective genes.

Gene therapy is already feasible. T. Friedmann, Science (1989) 244:1275-1281; M. Bluestone, Biotechnol (1992) 10:132-134. Systems and polymers for delivery of polynucleotides are known in the art. P.L. Felgner, Adv Drug Delivery Rev (1990) 5:163-187. Adenoviral vectors have been used to transfer CFTR to the cotton rat lung in vivo. M.A. Rosenfeld et al., Cell (1992) 68:143-155. Although high levels of transfection in vivo have been reported with the adenoviral vectors, non-viral delivery systems have a number of advantages and should be

vigorously developed. Rosenfeld et al., supra; M.A. Rosenfeld et al., Science (1991) 252:431-434.

During the past decade, a number of methods have been developed to introduce functional genes into mammalian cells in vitro. These techniques are applicable to gene therapy if the target cells can be removed from the body, treated, and the transfected cells amplified and then returned to the patient. This option is not possible for CF patients. At present the best in vivo transfection efficiencies are obtained with retroviruses (Bluestone, supra) and adenoviruses (Rosenfeld et al., supra). However the efficiency is variable and a concern is that virus based gene delivery might cause viral infection or cancer. Initial human clinical trials have revealed no acute complications of retroviral vectors but the possibility of long-term complications mandate careful patient monitoring. K. Cornetta et al., Human Gene Ther (1991) 2:3-14.

The risks of using viral based vectors and the conceptual advantages in using plasmid DNA constructs for gene therapy (discussed in P.L. Felgner et al., Nature (1991) 349:351-352) have led to a parallel development of various physical and chemical methods for gene transfer. The most intensely studied systems involve treatment of the cells with calcium phosphate or a cationic facilitator (Felgner et al., supra). Other popular methods involve DNA injection during physical puncture of the membrane (M.R. Capecchi, Cell (1980) 22:479-485) or passive uptake during permeabilization or abrasion of the cellular membrane (Felgner et al., supra). Each method is intrinsically aggressive and applicable only in vitro.

This invention is in the field of direct gene delivery that does not involve the use of viral vehicles.

A non-viral carrier for gene delivery must be able to surmount many barriers: it must survive in circulation, it must be able to target the cell of choice, it must be able to introduce DNA into the cytoplasm, and it must be able to transport the DNA into the nucleus.

Masking. One concern about the direct delivery of genes in vivo is the ability of the polynucleotide to survive in circulation long enough to arrive at the desired cellular destination. "Masking", or protecting the polynucleotides is one way to address this concern.

Microparticulates (such as the erythrocyte ghost, reconstituted viral envelopes and liposomes) have been used in part as protection in gene transfer. C. Nicolau et al., Crit Rev Ther Drug Carr Sys (1989) 6:239-271; R.J. Mannio et al., Biotechniques (1988) 6:682-690. The most successful liposome system uses the cationic lipid reagent dioleyloxytrimethylammonium (DOTMA). P.L. Felgner et al., Proc Natl Acad Sci (USA) (1987) 84:7413-7417. DOTMA is mixed with phosphatidylethanolamine (PE) to form the reagent Lipofectin™. The advantage of using Lipofectin™ is that the cationic liposome is simply mixed with the DNA and added to the cell. It is not necessary to encapsulate the DNA inside of the liposome with the cationic reagents. Lipofectin™ has been used to transfect reporter genes into human lung epithelial cells in culture (L. Lu et al., Pflugers Arch (1989) 415:198-203), to introduce the CAT gene into rats by intratracheal route (T.A. Hazinski et al., Am J Respir Cell Mol Biol (1991) 4:206-209) and to introduce the CAT gene into mice by the intratracheal and intravenous route (K.L. Brigham et al., Am J Med Sci (1989) 298:278-281; A. Bout et al., "Abstracts of the 1991 Cystic Fibrosis Conference", Abstract no. 87 (1991)). About 50% of the airway epithelial cells transiently expressed the ß galactosidase reporter gene (Hazinski et al., supra) but the level of expression was not quantitated. When chloramphenicol acetyltransferase (CAT) attached to a steroid sensitive promoter was transfected into rat lung, expression could be positively regulated by dexamethasone. Hazinski et al., supra. Cytotoxicity is a problem with high concentrations of Lipofectin™.

Substitutes for DOTMA include lipopolyamine (J. Loeffler et al., J Neurochem (1990) 54:1812-1815), lipophilic polylysines (X. Zhou et al., Biochim Biophys

Acta (1991) 1065:8-14) and a cationic cholesterol (X. Gao et al., Biochem Biophys Res Comm (1991) 179:280-285). These have been used to mediate gene transfer in culture. Although there is some improvement over transfection rates observed with Lipofectin™ (about threefold), toxicity remains a problem. Studies on the mechanism responsible for transfection using the cationic lipids have been notably lacking. The past approach has been to synthesize different cationic lipids and try them in transfection assays, rather than to systematically study how the delivery systems introduce DNA into the cell. DOTMA/PE liposomes can undergo bilayer fusion with anionic liposomes (N. Duzgunes et al., Biochem (1989) 28:9179-9184) which suggests that direct entry of the DNA via the plasma membrane is involved with DOTMA's mode of action. High efficiency transfection using cationic lipids systems requires the inclusion of PE, possibly because PE can form intramembrane lipid intermediates which facilitate membrane fusion. The role of PE in membrane permeabilization and fusion has been extensively studied. E.g., M.-Z. Lai et al., Biochem (1985) 24:1646-1653; H. Ellens et al., Biochem (1986) 25:285-294; J. Bentz et al., Biochem (1987) 26:2105-2116).

Cellular Targeting. Efficient gene transfer requires targeting of the DNA to the cell of choice. Recently, procedures based upon receptor mediated endocytosis have been described for gene transfer. G.Y. Wu et al., J Biol Chem (1987) 262:4429; G.Y. Wu et al., J Biol Chem (1988) 263:14621-14624. A cell-specific ligand-polylysine complex is bound to nucleic acids through charge interactions. The resulting complex is taken up by the target cells. Wu et al., supra, reported efficient transfection of the human hepatoma cell line HepG2 and of rat hepatocytes in vivo using this delivery system with asialoorosomucoid as a ligand. Huckett et al., Biochem Pharmacol (1990) 40:253-263, reported stable expression of an enzymatic activity in HepG2 cells following insulin-directed targeting. Finally Wagner et al., Proc Natl Acad Sci (USA) (1990) 87:3410-3414

and (1991) 88:4255-4259 observed transferrin-polycation-mediated delivery of a plasmid into the human leukemic cell line K-562 and subsequent expression of the encoded luciferase gene. However, the described delivery systems are based upon high molecular weight targeting proteins linked to DNA through a polylysine linker. These large ligand-polycation conjugates are heterogenous in size and composition, not chemically well-defined, and difficult to prepare in a reproducible fashion (Wu et al., supra; Wagner et al., supra). Moreover, in many of the receptor mediated systems, chloroquine or other disruptors of intracellular trafficking are required for high levels of transfection. In one study, adenovirus has been used to enhance gene delivery of the receptor mediated systems. D.T. Curiel et al., Proc Natl Acad Sci (USA) (1991) 88:8850-8854.

Together these studies show that genes can be delivered into the interior of mammalian cells by receptor mediated endocytosis and a fraction of the exogenous DNA escapes degradation, enters the nucleus, and is expressed. The level of expression is low, probably due to the limited entry of the foreign DNA into the cytoplasm.

Charge Neutralization and Membrane Permeabilization. Direct delivery of genes is aided by the ability to neutralize the large negative charge on the polynucleotide, and the (often concomitant) ability to permeabilize the membrane of the targeted cell. The use of polycations to neutralize the polynucleotide charge and aid in the membrane permeabilization and translocation is well known. Felgner, supra. Cationic lipids have also been used for this purpose. P.L. Felgner et al., Proc Natl Acad Sci (USA) (1987) 84:7413-7417; U.S. Patent No. 4,946,787 to Eppstein et al. Certain cationic lipids termed lipopolyamines and lipointercalants are also known. J.-P. Behr, Tet Lett (1986) 27:5861-5864.

Subcellular Localization. Once the polynucleotide has entered the targeted cell, direct delivery of genes would be aided by the ability to direct the genes to the proper subcellular location. One obvious target for the delivery

of deoxyribonucleotides is the nucleus. Ligands known to aid in this process are nuclear localization peptides, or proteins containing these nuclear localization sequences. C. Dingwall et al., TIBS (1991) 16:478-481.

Y. Kaneda et al., Science (1989) 243:375-378, showed that the transfection efficiency obtained with reconstituted viral envelopes is increased when the foreign gene is co-delivered into the target cells with nuclear proteins. DNA mixed with nuclear proteins exhibit a modest increase in transfection over DNA that was mixed with albumin (Kaneda et al., supra). The assumption is that the DNA is incorporated into the nucleus more readily when proteins containing the nuclear localization sequence (NLS) pro-lys-lys-lys-arg-lys-val (P.A. Silver, Cell (1991) 64:489-497) are associated with the plasmid. The NLS on a protein designates it for transport through the nuclear pore. Nuclear localization sequences of 14 amino acids have been attached to a variety of macromolecules and even gold particles (150 A diameter) and, when introduced into the cytoplasm, they are rapidly incorporated into the nucleus (D.R. Findlay et al., J Cell Sci Supp (1989) 11:225-242; Silver, supra). The suggestion that nuclear entry is rate limiting for successful, stable transfection is also supported by the finding that plasmid DNA microinjected into the cytoplasm is unable to bring about transfection of cells (no transfection out of 1000 cytoplasmic injections, whereas microinjection of plasmids directly into the nucleus results in transfection in greater than 50% of the microinjected cells. Cappechi, supra. If the attachment of nuclear localization signals on the plasmid leads to transport of plasmid DNA into the nucleus, the transfection efficiency should increase. We propose a novel method to attach NLS and other ligands to the desired polynucleotide.

Finally, investigators have demonstrated that transfection efficiencies increase when DNA is condensed using various cationic proteins. T.I. Tikchonenko et al., Gene (1988) 63:321-330; M. Bottger et al., Biochim Biophys

Acta (1988) 950:221-228; Wagner et al., supra. The reason why DNA condensation increases transfection is not readily apparent, it may increase cellular uptake of DNA (Wagner et al., supra) but it also can decrease susceptibility of the DNA to nuclease activity which may result in higher amounts of intact DNA in the cell.

Polynucleotide Association. Direct delivery of genes associated with one of the above-discussed classes of molecules, is further aided by the ability of those components to remain associated with the DNA. Wu et al., supra, associated their receptor ligand with the polynucleotide by covalently attaching the ligand to the polycation polylysine. Wagner et al., supra, in addition to polylysine, also covalently attached the ligand to a DNA intercalator, ethidium homodimer (5,5'-diazadecamethylene-bis(3,8-diamino-6-phenylphenanthridium) dichloride dihydrochloride). P.E. Nielsen, Eur J Biochem (1982) 122:283-289, associated photoaffinity labels to DNA by covalent attachment to 9-aminoacridine and certain bis-acridines.

None of the above references describe a system for polynucleotide delivery aimed at multiple aspects of the problems involved in bringing a circulating polynucleotide to a targeted subcellular location of a targeted cell. This invention addresses those problems by associating the polynucleotide with a combination of one or more of the following functional components: DNA-masking components, cell recognition components, charge-neutralization and membrane-permeabilization components, and subcellular localization components.

Summary of the Invention

In light of the aforementioned problems of direct gene delivery, this invention contemplates a self-assembling polynucleotide delivery system utilizing a combination of one or more, preferably two or more of the following functional components: DNA-masking components, cell recognition components, charge-neutralization and membrane-

permeabilization components, and subcellular localization components. Each component in this system is able to perform its indicated function and also be capable of assembling or disassembling with the polynucleotide as required. For example, certain components may have to dissociate from the polynucleotide in order for it to perform its desired function.

It is accordingly a primary object of this invention to provide a composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell comprising the polynucleotide associated with a membrane-permeabilizing component capable of transporting the polynucleotide across the cytoplasmic membrane of the eukaryotic cell.

It is another object of this invention to provide a composition for presenting a polynucleotide to the nucleus of a eukaryotic cell comprising the polynucleotide associated with a cell recognition component capable of recognizing the eukaryotic cell.

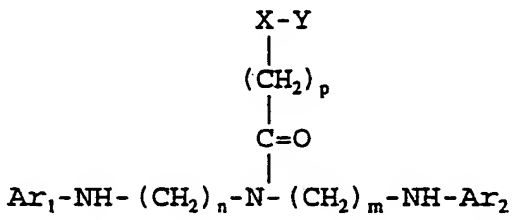
It is yet another object of this invention to provide a composition for presenting a polynucleotide to the nucleus of a eukaryotic cell comprising the polynucleotide associated with both a cell recognition component capable of recognizing the eukaryotic cell, and a membrane-permeabilizing component capable of transporting the polynucleotide across the cytoplasmic membrane of the eukaryotic cell.

It is a further object of this invention to provide a composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell comprising the polynucleotide associated with a subcellular-localization component capable of delivering the polynucleotide from the cytoplasm of the eukaryotic cell to a subcellular component of the eukaryotic cell.

It is still a further object of this invention to provide a composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell comprising the polynucleotide, a cell recognition component capable of recognizing said eukaryotic cell, a membrane-permeabilizing

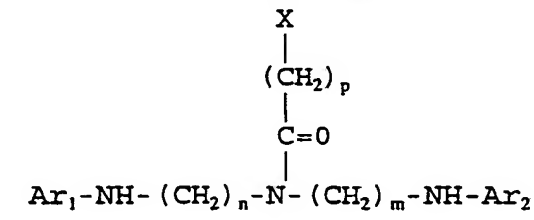
component capable of transporting the polynucleotide across the cytoplasmic membrane of said eukaryotic cell, a subcellular-localization component capable of delivering the polynucleotide from the cytoplasm of said eukaryotic cell to a subcellular component of said eukaryotic cell, and a masking component capable of increasing the circulatory half-life of the polynucleotide.

It is another object of this invention to provide a component useful in self-assembling polynucleotide delivery systems having the formula



wherein each of n and m is independently an integer of 1 to 20, p is an integer of 0 to 20, Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyrido-carbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof, X is a reactive coupling group, and Y is selected from the group consisting of cell surface receptor ligands, subcellular localization sequences, and membrane permeabilizing components.

It is still another object of this invention to provide a reactive intercalating component having the formula



wherein each of n and m is independently an integer of 1 to 20, p is an integer of 0 to 20, Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyrido-carbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof, and X is a reactive group.

Brief Description of the Figures

Figure 1 shows one embodiment of the polynucleotide delivery system of the invention, where NLS is a nuclear localization sequence, MD is a membran-permeabilization component, and Ligand is a cell recognition component.

Figure 2 shows the structure of gramicidin S.

Figure 3 compares the efficiency of luciferase transfection with Lipofectin™, pH-sensitive liposomes, and the gramicidin S/DOPE/DNA complex.

Figure 4 shows the effect of gramicidin S to DNA ratio on transfection efficiency.

Figure 5 shows the effect of gramicidin S to DOPE ratio on transfection efficiency.

Figure 6 shows the effect of lipid type in the gramicidin S/lipid/DNA complex on transfection efficiency.

Figure 7 shows the effect of substituting other peptides for gramicidin S in the gramicidin S/lipid/DNA complex on transfection efficiency.

Figure 8 shows a synthetic scheme for attaching targeting carbohydrates and/or reactive maleimide to spermidine bis-acridine.

Figure 9 shows the basic scheme for coupling peptides to the maleimido-spermidine bis-acridine.

Figure 10 shows a synthetic scheme for coupling to a degradable Lys-Lys peptide bis-acridine.

Figure 11 shows the results of the gel retardation assay of Example 3.

Figure 12 shows the ability of several galactosyl bis-acridines to bring plasmid DNA into hepatocytes.

Figure 13 shows a synthetic scheme for the trigalactosylated spermidine bis-acridine of Example 6.

Detailed Description of the Invention

Definitions:

The term "polynucleotide" as used herein, includes RNA or DNA sequences of more than one nucleotide in either single chain, duplex or multiple chain form. "Polynucleotide" is generic to polydeoxyribonucleotides

(containing 2'-deoxy-D-ribose or modified forms thereof), i.e., DNA, to polyribonucleotides (containing D-ribose or modified forms thereof), i.e., RNA, and to any other type of polynucleotide which is an N-glycoside or C-glycoside of a purine or pyrimidine base, or modified purine or pyrimidine base or abasic nucleotides. The polynucleotide may encode promoter regions, operator regions, structural regions, termination regions, combinations thereof or any other genetically relevant material.

The polynucleotides of the invention may also contain one or more "substitute" linkages as is generally understood in the art. Some of these substitute linkages are non-polar and contribute to the desired ability of the polynucleotide to diffuse across membranes. Others contribute to the increased or decreased biodegradability of the polynucleotide. (Biodegradability will be affected, for example, by increased or decreased nuclease sensitivity.) These "substitute" linkages are defined herein as conventional alternative linkages such as phosphorothioate or phosphoramidate, are synthesized as described in the generally available literature. Not all such linkages in the same polynucleotide need be identical.

Modifications in the sugar moiety of the polynucleotide, for example, wherein one or more of the hydroxyl groups are replaced with halogen, aliphatic groups, or functionalized as ethers, amines, and the like, or wherein the ribose or deoxyribose is replaced with other functionally equivalent structures, are also included. Modifications in the base moiety include alkylated purines or pyrimidines, acylated purines or pyrimidines, or other heterocycles. Such "analogous purines" and "analogous pyrimidines" are those generally known in the art, many of which are used as chemotherapeutic agents.

In particular, the sugar-phosphate backbone of the polynucleotide may be replaced with a non-carbohydrate backbone such as a peptide or other type of polymer backbone as discussed in P.E. Nielsen et al., Science (1991) 254:1497-1500.

The term "functional component" as used herein, includes DNA-masking components, cell recognition components, charge-neutralization and membrane-permeabilization components, and subcellular-localization components.

The term "DNA-masking component", as used herein, refers to a molecule capable of masking all or part of the polynucleotide, thereby increasing its circulatory half-life by inhibiting attack by degrading reagents (such as nucleases) present in circulation.

The term "membrane-permeabilizing component", as used herein, refers to any component that aids in the passage of a polynucleotide across a membrane. Thus, this term encompasses in part charge-neutralizing components, usually polycations, that neutralize the large negative charge on polynucleotides, and enable the polynucleotide to transverse the hydrophobic interior of a membrane. Many charge-neutralizing components can act as membrane-permeabilizers. Membrane-permeabilization may also arise from amphipathic molecules.

A membrane permeabilizer is a molecule that can assist a normally impermeable molecule to traverse cellular membranes and gain entrance to the cytoplasm of the cell. A membrane permeabilizer may be a peptide, bile salt, glycolipid, carbohydrate, phospholipid or detergent molecule. Membrane permeabilizers often have amphipathic properties such that one portion is hydrophobic and another is hydrophilic, permitting them to interact with membranes.

The term "liposome" as used herein, refers to small vesicles composed of amphipathic lipids arranged in spherical bilayers. Liposomes are usually classified as small unilamellar vesicles (SUV), large unilamellar vesicles (LUV), or multi-lamellar vesicles (MLV). SUVs and LUVs, by definition, have only one bilayer, whereas MLVs contain many concentric bilayers. Liposomes may be used to encapsulate various materials, by trapping hydrophilic molecules in the aqueous interior or between bilayers, or by trapping hydrophobic molecules within the bilayer.

Liposomes exhibit a wide variety of characteristics, depending upon their size, composition, and charge. For example, liposomes having a small percentage of unsaturated lipids tend to be slightly more permeable, while liposomes incorporating cholesterol or other sterols tend to be more rigid and less permeable. Liposomes may be positive, negative, or neutral in charge, depending on the hydrophilic group. For example, choline-based lipids impart an overall neutral charge, phosphate and sulfate based lipids contribute a negative charge, glycerol-based lipids are generally negatively-charged, and sterols are generally neutral in solution but have charged groups.

The term "cell recognition component" as used herein, refers to a molecule capable of recognizing a component on the surface of a targeted cell. Cell recognition components include: antibodies to cell surface antigens, ligands for cell surface receptors including those involved in receptor-mediated endocytosis, peptide hormones, and the like.

The term "DNA-associating moiety" refers to a molecule or portions thereof that interacts in a noncovalent fashion with nucleic acids. DNA-associating moieties include major- and minor-groove binders, which are molecules thought to interact with DNA by associating with the major or minor groove of double-stranded DNA. DNA associating moieties also include DNA intercalators, which are planar molecules or planar portions of molecules thought to intercalate into DNA by inserting between and parallel to nucleotide base pairs. DNA associating moieties further include polycations, thought to associate with the negative charges on the DNA backbone. When a single-stranded DNA or RNA is used as the therapeutic strand, the complementary "linker strand" as described herein may functionally act as the "DNA-associating moiety".

DNA associating moieties may be covalently linked through a "reactive group" to a functional component of this invention. These reactive groups are easily reacted with a nucleophile on the functional component. Such

reactive groups (with their corresponding reactive nucleophiles) include, but are not limited to: N-hydroxysuccinimide (amine), maleimide and maleimidophenyl (sulfhydryl), pyridyl disulfide (sulfhydryl), hydrazide (carbohydrate), and phenylglyoxal (arginine).

The term "subcellular-localization component" as used herein, refers to a molecule capable of recognizing a subcellular component in a targeted cell. Recognized subcellular components include the nucleus, ribosomes, mitochondria, and chloroplasts. Particular subcellular-localization components include the "nuclear-localization components" that aid in bringing molecules into the nucleus and are known to include the nuclear localization peptides and amino acid sequences.

The Compositions:

The compositions of this invention in part are self-assembling polynucleotide delivery systems utilizing a polynucleotide in combination with one or more, preferably two or more of the following functional components: DNA-masking components, cell recognition components, charge-neutralization and membrane-permeabilization components, and subcellular localization components. Each element in this system is able to perform its indicated function and also be capable of assembling or disassembling with the polynucleotide as required. Individual elements of this system, and methods and intermediates for making these elements are also contemplated as part of this invention. One embodiment of the system is shown in Figure 1.

The Polynucleotide

The polynucleotide may be single-stranded DNA or RNA, or double-stranded DNA or DNA-RNA hybrid. Triple- or quadruple-stranded polynucleotides with therapeutic value are also contemplated to be within the scope of this invention. Examples of double-stranded DNA would include structural genes, genes including operator control and

termination regions, and self-replicating systems such as plasmid DNA.

Single-stranded polynucleotides include antisense polynucleotides (DNA and RNA), ribozymes and triplex-forming oligonucleotides. This "therapeutic strand", in order to have prolonged activity, preferably has as some or all of the nucleotide linkages stable, non-phosphodiester linkages. Such linkages include, for example, the phosphorothioate, phosphorodithioate, phosphoroselenate, or O-alkyl phosphotriester linkages wherein the alkyl group is methyl or ethyl.

For these single-stranded polynucleotides, it may be preferable to prepare the complementary strand to the therapeutic strand as part of the administered composition. This complementary strand is designated the "linker strand", and is usually synthesized with a phosphodiester linkage so that it is degraded after entering the cell. The "linker strand" may be a separate strand, or it may be covalently attached to or a mere extension of the therapeutic strand so that the therapeutic strand essentially doubles back and hybridizes to itself.

The linker strand may also have functionalities on the 3' or 5' end or on the carbohydrate or backbone of the linker that serve as functional components to enhance the activity of the therapeutic strand. For example, the phosphodiester linker strand may contain a targeting ligand such as a folate derivative that permits recognition and internalization into the target cells. If the linker is attached to its complementary therapeutic strand that is composed of degradation-resistant linkages, the duplex would be internalized. Once inside the cell, the linker would be degraded, releasing the therapeutic strand. In this manner the therapeutic strand would have no additional functionalites attached and its function would not be impeded by non-essential moieties. This strategy could be applied to any antisense, ribozyme or triplex-forming polynucleotide. It would be used to deliver antiviral, antibacterial, antineoplastic, antiinflammatory,

antiproliferative, anti-receptor blocking or anti-transport polynucleotides and the like.

A separate "linker strand" may be synthesized to have the direct complementary sequence to the therapeutic strand and hybridize in a one-on-one fashion. Alternatively, the linker strand may be constructed so that the 5′ region of the linker strand hybridizes to the 5′ region of the therapeutic strand, and the 3′ region of the linker strand hybridizes to the 3′ region of the therapeutic strand to form a concatenate of the following structure

5′ _____ _____ _____

3′ _____ _____ _____

This concatenate has the advantage that the apparent molecular weight of the therapeutic nucleic acids is increased and its pharmacokinetic properties and targeting ligand:therapeutic ligand ratio can be adjusted to achieve the optimal therapeutic effect.

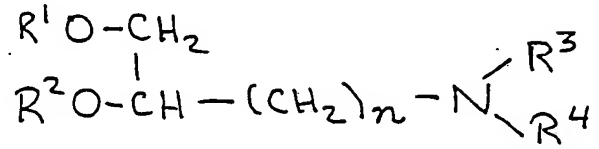
The Functional Components

DNA-Masking Components. The DNA-masking element of this system is a molecule capable of masking all or part of the polynucleotide, thereby increasing its circulatory half-life by inhibiting attack by degrading reagents present in circulation.

In this invention, polyethylene glycol (PEG) can be covalently linked with a DNA-associating moiety by conventional methods as described below, and used as a DNA-masking component. The PEG will have a molecular weight from about 700 to about 20,000 daltons, preferably about 1800 to 6000 daltons, and is preferably present in a ratio (molecules PEG:bp DNA) from about 1:4 to 1:100, more preferably about 1:20.

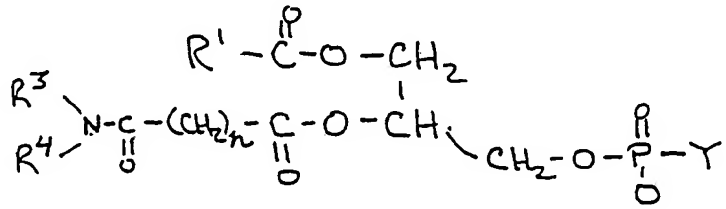
Alternatively, DNA may be masked through association with lipids. In one embodiment, the DNA is encased in standard liposomes as described, for example, in U.S. Patent No. 4,394,448 to Szoka et al., the specification of which is hereby incorporated by reference. In another embodiment, the DNA is incubated with a synthetic cationic

lipid similar to those described in U.S. Patent No. 4,897,355 to Eppstein et al. These cationic lipids have the general formula



wherein n is an integer from 1 to 8, R^1 and R^2 are the same or different and are alkyl or alkenyl having from 6 to 24 carbon atoms, R^3 is hydrogen, alkyl or alkylamine having from 1 to 10 carbon atoms, and R^4 is a positively charged linear or branched alkyl or alkylamine having from 1 to 30 carbon atoms, wherein one or more of the carbon atoms may be substituted with NR', wherein R' is hydrogen, alkyl or alkylamine having from 1 to 10 carbons. Preferred groups that can function as the -N-R' moiety are tris(aminoethyl)amine $(NH_2CH_2CH_2)_3N$, agmatine (decarboxyarginine) $H_2N(CH_2)_4C(=NH)NH_2$, 3-aminoethyl-1,3-propanediamine $H_2N(CH_2)_3NH(CH_2)_2NH_2$, 3-dimethylaminopropylamine $(CH_3)_2NH(CH_2)_3NH_2$, iminobis(N,N')dimethylpropylamine $NH((CH_2)_3N(CH_3)_2)_2$, iminobis(3-aminopropyl)-1,3-propanediamine, 1,4-bis(3-aminopropyl)piperazine, bis(propylamine) $(NH_2(CH_2)_3)_2NH$, spermidine, and spermine, wherein these groups are attached to the lipid molecule through one of their nitrogen atoms.

In a specifically preferred embodiment, the synthetic cationic lipid is a synthetic cationic tail lipid having the formula



wherein n is an integer from 6 to 24, Y is selected from the group consisting of hydrogen, ethanolamine, choline, glycerol, serine and inositol, R^1 is alkyl or alkenyl having from 6 to 24 carbon atoms, R^3 is hydrogen, alkyl or alkylamine having from 1 to 10 carbon atoms, and R^4 is a

positively charged linear or branched alkyl or alkylamine having from 1 to 30 carbon atoms, wherein one or more of the carbon atoms may be substituted with NR', wherein R' is hydrogen, alkyl or alkylamine having from 1 to 10 carbons. Preferred groups that can function as the -N-R' moiety are tris(aminoethyl)amine $(NH_2CH_2CH_2)_3N$, agmatine (decarboxyarginine) $H_2N(CH_2)_4C(=NH)NH_2$, 3-aminoethyl-1,3-propanediamine $H_2N(CH_2)_3NH(CH_2)_2NH_2$, 3-dimethylaminopropylamine $(CH_3)_2NH(CH_2)_3NH_2$, iminobis(N,N')dimethylpropylamine $NH((CH_2)_3N(CH_3)_2)_2$, iminobis(3-aminopropyl)-1,3-propanediamine, 1,4-bis(3-aminopropyl)piperazine, bis(propylamine) $(NH_2(CH_2)_3)_2NH$, spermidine, and spermine, wherein these groups are attached to the lipid molecule through one of their nitrogen atoms.

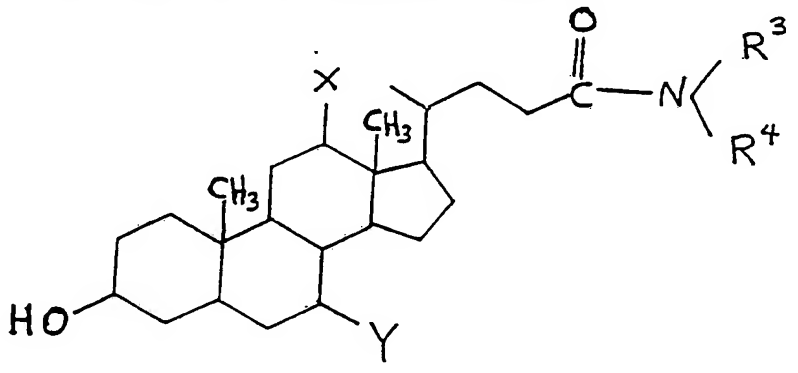
It has been found that the above-described synthetic cationic lipids effectively mask the DNA when associated therewith. Without attempting to limit the invention in any way, it is believed that the lipids form a monolayer structure that encapsulates the DNA in some fashion.

Cell Recognition Components. The cell recognition element of this system is a molecule capable of recognizing a component on the surface of a targeted cell, covalently linked with a DNA-associating moiety by conventional methods as described below. Cell recognition components include: antibodies to cell surface antigens, ligands for cell surface receptors including those involved in receptor-mediated endocytosis, peptide hormones, etc. Specific ligands contemplated by this invention include: carbohydrate ligands such as galactose, mannose, mannosyl 5-phosphate, fucose, sialic groups, N-acetylglucosamine or combinations of these groups as complex carbohydrates such as those found on glycolipids of the blood groups or on various secreted proteins. Other ligands include folate, biotin, various peptides that can interact with cell surface or intracellular receptors such as the chemoattractant peptide N-formyl-met-leu-phe, peptides containing the arg-asp-glycine sequence or cys-ser-gly-arg-glu-asp-val-trp peptides, peptides that contain a cystine

residue or that interact with cell surface protein such as the human immunodeficiency virus GP-120, and peptides that interact with CD-4. Other ligands include antibodies or antibody fragments such as described by A. Hertler and A. Frankel, J Clin Oncol 7: 1932-1942. The specificity of the antibodies can be directed against a variety of epitopes that can be expressed on cell surfaces including histocompatability macromolecules, autoimmune antigens, viral, parasitic or bacterial proteins. Other protein ligands include hormones such as growth hormone and insulin or protein growth factors such as GM-CSF, G-CSF, erythropoetin, epidermal growth factor, basic and acidic fibroblast growth factor and the like. Other protein ligands would include various cytokines that work through cell surface receptors such as interleukin 2, interleukin 1, tumor necrosis factor and suitable peptide fragments from such macromolecules.

Membrane-Permeabilizing Components. The membrane-permeabilizing element of this system is a molecule that aids in the passage of a polynucleotide across a membrane. The liposomes and synthetic cationic lipids described above as DNA-masking components also may function as membrane-permeabilization components.

The membrane-permeabilizing components of this invention also include polycations that neutralize the large negative charge on polynucleotides. Polycations of this invention include polylysine, polyarginine, poly (lysine-arginine) and similar polypeptides, and the polyamines. Another class of polycations are the cationic bile salts having the following formula:



wherein X and Y are independently H or OH, R^3 is hydrogen, alkyl or alkylamine having from 1 to 10 carbon atoms, and R^4 is a positively charged linear or branched alkyl or alkylamine having from 1 to 30 carbon atoms, wherein one or more of the carbon atoms may be substituted with NR', wherein R' is hydrogen, alkyl or alkylamine having from 1 to 10 carbons. Preferred groups that can function as the -N-R' moiety are tris(aminoethyl)amine $(NH_2CH_2CH_2)_3N$, agmatine (decarboxyarginine) $H_2N(CH_2)_4C(=NH)NH_2$, 3-aminoethyl-1,3-propanediamine $H_2N(CH_2)_3NH(CH_2)_2NH_2$, 3-dimethylaminopropylamine $(CH_3)_2NH(CH_2)_3NH_2$, iminobis(N,N')dimethylpropylamine $NH((CH_2)_3N(CH_3)_2)_2$, iminobis(3-aminopropyl)-1,3-propanediamine, 1,4-bis(3-aminopropyl)piperazine, bis(propylamine) $(NH_2(CH_2)_3)_2NH$, spermidine, and spermine, wherein these groups are attached to the bile salt through one of their nitrogen atoms.

In a different embodiment, the membrane-permeabilizing component of the invention is an amphipathic cationic peptide. Amphipathic cationic peptides are peptides whose native configuration is such that the peptide is considered to have a cationic face and a neutral, hydrophobic face. In a preferred embodiment, the peptide is a cyclic peptide. Examples of the amphipathic cationic cyclic peptides of this invention are gramicidin S (the structure of which is shown in Figure 2), and the tyrocidines. The peptide may also contain some or all of the amino acids in the D configuration as opposed to the naturally occurring L configuration.

In a particularly preferred embodiment, the membrane-permeabilizing element includes, in addition to the amphipathic cationic cyclic peptides, either (1) a lipid, or (2) a simple polyamine, or both.

The lipid of the invention is an amphipathic molecule which is capable of liposome formation, and is substantially non-toxic when administered at the necessary concentrations either in native form or as liposomes. Suitable lipids generally have a polar or hydrophilic end, and a non-polar or hydrophobic end. Suitable lipids

include without limitation egg phosphatidylcholine (EPC), phosphatidylethanolamine, dipalmitoylphosphatidylcholine (DPPC), cholesterol (Chol), cholesterylphosphorylcholine, 3,6,9-trioxaoctan-1-ol-cholesteryl-3e-ol, dimyristoyl-phosphatidylcholine (DMPC), and other hydroxy-cholesterol or aminocholesterol derivatives (see, e.g., K.R. Patel et al., Biochim Biophys Acta (1985) 814:256-64). The lipid is preferably added in the form of liposomes.

The added polyamine is preferably spermine or spermidine.

The membrane permeabilizing elements -- the cyclic peptide and optional phospholipid and polyamine -- may be added to the composition simultaneously or consecutively. Preferably, the cyclic peptide is added first, and the phospholipid or polyamine added later. The molar ratio of added cyclic peptide to added polyamine is preferably from about 1:1 to about 1:3. The molar ratio of added cyclic peptide to added phospholipid is preferably from about 1:1 to about 1:20.

Subcellular-Localization Components. The subcellular-localization element of this system is a molecule capable of recognizing a subcellular component in a targeted cell, covalently linked with a DNA-associating moiety by conventional methods as described below. Particular subcellular components include the nucleus, ribosomes, mitochondria, and chloroplasts.

In a preferred embodiment of this invention, the subcellular-localization component is a nuclear-localization component. The nuclear-localization components include known peptides of defined amino acid sequences, and longer sequences containing these peptides. One known peptide sequence is the SV 40 large T antigen heptapeptide pro-lys-lys-lys-arg-lys-val. Other peptides include the influenza virus nucleoprotein decapeptide ala-ala-phe-glu-asp-leu-arg-val-leu-ser, and the adenovirus E1a protein sequence lys-arg-pro-arg-pro. Other sequences may be discerned from C. Dingwall et al., TIBS (1991) 16:478-481.

In another embodiment, the subcellular-localization component is a lysosomal-localization component. A known component for targeting the lysosome is a peptide containing the sequence lys-phe-glu-arg-gln. In yet another embodiment, the subcellular-localization component is a mitochondrial-localization component. A known component for targeting mitochondria is a peptide containing the sequence met-leu-ser-leu-arg-gln-ser-ile-arg-phe-phe-lys-pro-ala-thr-arg.

DNA-Associating Moieties

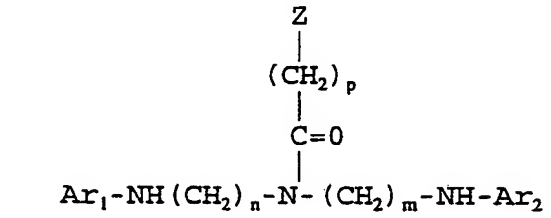
The DNA-associating moiety of this system refers to a portion of a functional component that interacts in a noncovalent fashion with nucleic acids. The moiety is covalently linked to the rest of the functional component by conventional means or as described below. DNA-associating moieties are preferably major- and minor-groove binders, DNA intercalators, or general DNA binders. In the case of single-stranded polynucleotides, the DNA-associating moiety may even be the linker strand as described above. In such a case the functional moiety, such as the cell-recognition or subcellular-localization component is covalently linked to the linker strand.

In one preferred embodiment, the DNA-associating moiety is a major- or minor-groove binder. The major- and minor-groove binders are moieties known to associate or "lay in" the major or minor groove of DNA. These binders include distamycin A and Hoechst dye 33258.

In another embodiment, the DNA-associating moiety is a nonspecific DNA binder such as a polycation. Polycations of this invention include polylysine, polyarginine, poly (lysine-arginine) and similar polypeptides, and the polyamines.

In another preferred embodiment, the DNA-associating moiety is a DNA intercalator. DNA intercalators are planar polycyclic molecules such as ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof. In a particular preferred embodiment, the intercalator is a

dimer consisting of two covalently linked planar polycyclic molecules. A planar polycyclic dimer moiety of this invention has the structure



wherein Z is a bond, each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20; and

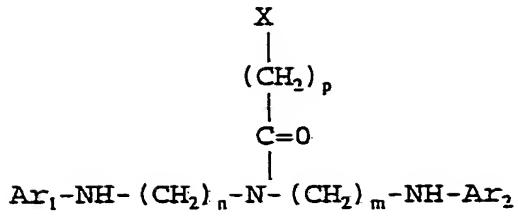
Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof.

The values of n and m are important as they determine the spacing of the intercalated acridine monomers in the DNA. More preferred values of n and m are 3 and 4, respectively. Bis-acridine dimers, wherein Ar_1 and Ar_2 are both acridine, are preferred.

This preferred DNA-associating moiety will be covalently attached to a functional moiety, said moiety being a cell recognition moiety, subcellular localization moiety, or membrane permeabilizing moiety as described above. The value of p determines the separation of the intercalator from the functional moiety. Preferred values for p are from 0 to 8.

The DNA-associating moiety may be covalently attached to multiple copies of, or more than one functional moiety. For example, the bis-acridine dimer may be attached to three galactose residues that bind to the hepatocyte asialoorosomucoid receptor as shown in Figure 13.

A preferred method for attaching the DNA-associating dimer to the functional moiety involves a precursor having the formula

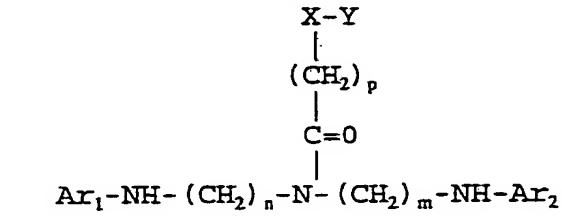


wherein each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20;

Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof; and

X is a reactive group selected from the group consisting of N-hydroxysuccinimide, maleimide, maleimidophenyl, pyridyl disulfide, hydrazide, and phenylglyoxal.

In a preferred embodiment, Ar_1 and Ar_2 are acridine, p is 4 and X is *p*-maleimidophenyl. This intercalating moiety is then coupled to the functional moiety through a sulfhydryl group on the functional moiety, for example, to obtain a bifunctional component having the structure



wherein Y is a functional component;

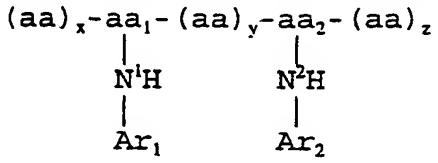
each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20;

Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof; and

X is a reactive group selected from the group consisting of N-hydroxysuccinimide, maleimide, maleimidophenyl, pyridyl disulfide, hydrazide, and phenylglyoxal.

Biodegradable linkers such as peptides having the sequence -lys-lys- may also be used in attaching the functional component to the intercalator.

In yet another embodiment of this invention, the planar polycyclic dimer has the formula



wherein Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof;

each aa is independently an amino acid;

x and z are integers independently selected from 1 to 100;

y is an integer from 0 to 5;

aa_1 and aa_2 are lysine residues;

N^1 and N^2 are nitrogens from the ϵ-amino groups of lysine residues aa_1 and aa_2.

Utility of the Polynucleotide Delivery System

The polynucleotide delivery system of the invention is useful in a therapeutic context. In therapeutic applications, the system of the invention can be formulated for a variety of modes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA, latest edition.

For systemic administration, parenteral administration such as injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For treating disorders of the lung, administration of the polynucleotide delivery system is done by inhalation or installation of the system directly into the lung.

For injection, the systems of the invention are formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the systems may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the systems can be administered orally, or through intranasal or inhaled aerosols. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays, for example, or using suppositories. For oral administration, the systems are formulated into conventional oral administration forms such as capsules, tablets, and tonics.

For topical administration, the systems of the invention are formulated into ointments, salves, gels, or creams, as is generally known in the art.

The following examples are meant to illustrate, but not to limit the invention.

Example 1

Gramicidin S Transfection

Lipofectin™ is a synthetic cationic lipid, dioleyloxytrimethylammonium, (DOTMA) in combination with phosphatidylethanolamine to form a charge complex with the negatively charged DNA. This complex is thought to fuse with the cell membrane and deliver DNA into the cytoplasm. An alternative approach uses pH sensitive liposomes composed of a negatively charged lipid and phosphatidylethanolamine. C.Y. Wang et al., Biochem (1989) 28:9508-9514. The delivery mechanism involves endocytosis

of the liposome, as the pH in the endosome becomes acidic, the liposomal bilayer destabilizes and fuses with the endosomal membrane. The contents of the liposome are then introduced into the cytoplasm of the cell. C.-J. Chu et al., Pharmaceut Res (1990) 7:824-834.

We have compared Lipofectin™ to a pH-sensitive cholesterylhemisuccinate (Chems)/phosphatidylethanolamine (PE) liposome composition and to gramicidin S / dioleoyl-phosphatidylethanolamine (DOPE) / DNA complexes for the delivery and expression of DNA in mammalian cells. Plasmids containing strong promoters and either firefly luciferase or β galactosidase were used as indicators for gene transfer.

Cell transfection protocol.

CV-1, p388D1, HepG2 and HeLa cells were provided by the UCSF Cell Culture Facility. The Lipofectin™ reagent was used as described in the product insert (Gibco-BRL, Gaithersburg, MD). KD83 cells were obtained from DNAX (Palo Alto, CA). Cells were plated at a density of 0.5-1X10^6 cells per 60 mm dish and grown 16 to 20 hrs at 37C under 5% CO_2 in appropriate media containing 10% fetal calf serum (FCS). Prior to incubation either with liposomes, Lipofectin™, or the gramicidin S/DOPE/DNA complex, cells were washed once with 2 ml of FCS-free DME H-21 medium. The transfection system was then added in 2 ml of the same media. In some experiments, transfection took place in 10% FCS containing DME H-21. After 5 hrs. media was removed and replaced by 3 ml of appropriate media with 10% FCS. Luciferase activity was measured after 48 hrs as described (A.R. Brasier et al., Biotechniques (1989) 7:1116-1122). Briefly, cells were washed twice with ice-cold phosphate buffer saline without Ca^{2+} and Mg^{2+} (PBS), treated with 400 µl of 25 mM glycylglycine (pH 7.8) in lysis buffer (containing 1% Triton) and scraped. After centrifugation, 100 µl of supernatant were mixed with an optimal amount of 50 mM ATP. D-luciferin (Sigma, 100 µl of a 1mM solution) was then injected and the emitted light was integrated during the first 10 sec. using a bioluminometer

(Bioluminescence Analytical Laboratories Inc., San Diego, CA). Proteins in the supernatant were assayed using the technique of Bradford (Bio-Rad kit). Results were expressed as light units per mg of cell-protein.

Luciferase Assay

In order to compare the potency of three different viral luciferase gene promoters, RSV, SV40 and CMV, we have transfected several mammalian cell lines with the corresponding Lipofectin™ complexed-plasmids. Each dish of cells received 2 μl of plasmid combined with 10 μl of Lipofectin™ as described above. Promoter strength was estimated by the luciferase expression at 48 hr given by the corresponding plasmid. The CMV promoter (pCluc4 plasmid) led to the highest luciferase expression in HeLa, HepG2 and p388D1 cells, while SV40 promoter (pSV2 plasmid) was more potent in CV-1 cells. Therefore for further experiments, pSV2 plasmid has been used in CV-1 cells and pCluc4 in other cell-lines.

Liposome characterization

Plasmid encapsulation efficiency was determined after separation of encapsulated from non-encapsulated plasmid on Ficoll gradients. About 22 ± 3% of the total DNA added was encapsulated. Liposome diameter was measured by dynamic light scattering and were 372±38 nm, 295±65 nm and 464±20 nm for DOPE/CHEMS, DOPC/CHEMS and PS/Chol liposomes respectively (results are the mean±SD of three independent light scattering determinations).

A. Gramicidin S and Phosphatidylethanolamine

Typical complex preparation was made by diluting 20 μg of plasmid DNA in 300 μl of 30 mM Tris Cl pH 9 in a polystyrene tube. Gramicidin S was diluted in 30 mM pH 9 Tris Cl buffer to a concentration of 2 mg/ml from a stock solution at 20 mg/ml in DMSO. 20 μl of diluted gramicidin S (i.e. 40 μg) solution was added to the DNA and quickly mixed. Then 170 nmoles of liposomes were added slowly drop by drop to the DNA gramicidin S mixture. Liposomes were prepared by drying 4 μmoles of lipids under nitrogen with

a rotavapor and by rehydrating the film with 4 ml of 30 mM pH 9 Tris Cl buffer. Liposomes were subsequently sonicated 30 min under argon using a bath sonicator. The diameter of the complex was determined by dynamic light scattering. Other peptides including, tyrocidine (U.S. Biochemicals), polymyxin B (Sigma) and polylysine 100 (Sigma) were also used to form the complex with DNA and lipids.

The efficiency of transfection was monitored by measuring the expression of luciferase in CV-1 cells as described above. The dose response comparing the amount of DNA added in the three transfection systems is illustrated in Figure 3. Light units per mg cell protein in a log scale are plotted on the Y axis and the amount of DNA added on the X axis. The open box designate results using the Gramicidin S-dioleoylphosphatidyl ethanolamine-DNA complex. This complex induces a 10 fold greater level of expression than obtained with Lipofectin, and a 1000 to 10,000 fold greater level of expression than obtained using the pH sensitive liposomes.

B. Gramicidin S-DNA Ratio Effects

The gramicidin S-DOPE-DNA complex was prepared as described in Example 1-A except the amount of gramicidin S added to the complex was varied at constant amounts of DNA (20 ug) and DOPE (170 nmoles). The complex was added to CV-1 cells and the luciferase activity measured as described in Example 1. The result is presented in Figure 4 and illustrates that maximum expression using the gramicidin S-DOPE-DNA complex occurs when the charge on the DNA is neutralized by the charge on the gramicidin.

C. Lipid Concentration Effects

The gramicidin S-DOPE-DNA complex was prepared as described in Example 1 except the amount of DOPE added to the complex was varied at constant amounts of DNA (20 ug) and gramicidin S (40 μg). The complex was added to CV-1 cells and the luciferase activity measured as described in Example 1. The result is presented in Figure 5, which illustrates that in the absence of the DOPE, expression is low. Maximum expression using the gramicidin S-DOPE-DNA

complex occurs when the ratio of DOPE to gramicidin S is above 5/1:mole/mole.

D. Lipid Type Effects

The gramicidin S-lipid-DNA complex was prepared as described in Example 1 except the type of phospholipid added to the complex was varied at constant amounts of DNA (20 ug) and gramicidin S(40 ug). The lipid compositions employed were DOPE; DOPE: dioleoylphosphatidylcholine (DOPC):2/1, palmitoyloleoylphosphatidylethanolamine (POPE), monomethyl DOPE (mmDOPE); dimethyl DOPE (dm DOPE); DOPC and dipalmitoylphosphatidylethanolamine (DPPE). The complex was added to CV-1 cells and the luciferase activity measured as described in Example 1. The result is presented in Figure 6, which illustrates that expression of luciferase activity is maximal with DOPE or a mixture of DOPE/DOPC:2/1 in the complex. Luciferase activity is appreciably diminished when the amino group on the DOPE is substituted with 2 (dm DOPE) or 3 methyl groups (DOPC). Expression of the encoded gene is also appreciably reduced when DPPE is used. This latter lipid has saturated acyl chains and a high transition temperature, which means the acyl chains of DPPE are less fluid than the other lipids tested in this series.

E. Effects of Added Non-Amphipathic Positively Charged Spermidine

The data presented in Example 2 show that gene expression due to the gramicidin S-DOPE-DNA complex is maximal when the negative charges on DNA are neutralized by the positive charges on gramicidin. To determine whether charge neutralization or membrane permeabilization is more important for gene transfer using this system, the positive charge contribution from gramicidin S was incrementally replaced by the positively charged polyamine, spermidine. The gramicidin S-lipid-DNA complex was prepared as described in Example 1 except the amount of gramicidin S added to the complex was varied at constant amounts of DNA (20 ug). The requisite positive charges required to neutralize the DNA was supplied by spermidine. The complex

was prepared with or without 170 nmoles of DOPE. The complex was added to CV-1 cells and the luciferase activity measured as described in Example 1. The results are given in Table 1 below, with luciferase activity expressed as light units/mg cell protein. First activity was always greater when DOPE was present in the complex. In the absence of DOPE, sequential replacement of positive charge due to gramicidin S by spermidine leads to a biphasic response; the expression of luciferase initially increase to a value about 100 fold less than the maximal response obtained in the presence of DOPE. When the percent of charge neutralization due to gramicidin S dropped below 25% transfection activity was totally lost. Thus, membrane permeabilization function of gramicidin S is more important than the charge neutralization function.

Table 1

Spermidine Charge Neutralization

% charges brought by GS	w/o lipids	with lipids
100	$4.5 \pm 2\ 10^3$	$8.5 \pm 0.7\ 10^8$
75	$4 \pm 2.5\ 10^5$	$5 \pm 2\ 10^8$
25	$2 \pm 2.5\ 10^6$	$2 \pm 0.5\ 10^7$
12.5	0	$2 \pm 0.5\ 10^7$

F. Use of Other Positively Charged Peptides

The peptide-DOPE-DNA complex was prepared as described in Example 1 except the type of peptide added to the complex was varied at constant amounts of DNA (20 ug) and DOPE (170 nmoles). The peptides employed were

polymixin B, a cyclic cationic peptide; polylysine, a linear cationic peptide; tyrocidine, a cyclic cationic peptide with a similar structure to gramicidin S but containing only a single positive charge and gramicidin S. The luciferase plasmid was also transfected into the cells using Lipofectin. The complex was added to CV-1 cells and the luciferase activity measured as described in Example 1. Figure 7 shows that gramicidin S induced the greatest level of expression followed closely by the related cyclic peptide tyrocidine. Both cyclic peptides were superior to Lipofectin at transferring the DNA into cells. Activity was also seen with the other two peptides, polymixin B and polylysine, however, the level of luciferase expression mediated by these two cationic peptides was inferior to that induced by gramicidin S or tyrocidine.

G. Comparison of Transfection Mediated by a DNA-Dendrimer Complex to That Obtained with a DNA-Polylysine Complex

To find better chemically-defined alternatives to the polyamine polymers such a polylysine, we have employed the hydrophyllic branched polycation macromolecules also known as the Starburst™ Dendrimer microparticles, Tomalia et al., supra, to form a complex with DNA or with DNA and the permeabilizing amphipathic peptide GALA. R. Parente et al., Bichemistry (1990) 29:8720-8728. The complex was prepared by diluting 12 μg of pCLUC4 plasmid in 660 μl of HBS (20 mM Hepes, 150 mM NaCl, pH 7.4) in a polystyrene tube. Polylysine (Sigma Chemical Co.) or Starburst™ Dendrimer microparticles of the fifth generation (1 nmole) (Polysciences, Inc.) was dissolved in 340 μl of HBS and added slowly (dropwise) to the DNA solution. In these conditions, the positive charges from the epsilon amino groups of the polylysines or from the peripheral amines of the dendrimers are in 1.3-fold excess over the negative charges of the plasmids. When the peptide GALA wa added, it was added so that the negative charges on GALA neutralized the excess charges on the dendrimer. The mixture was left to stand for thirty minutes after the last

addition at room temperature and then 500 μl of the mixture was added to CV-1 cells. The transfection protocol was carried out as described above. In this experiment, the best transfection protocol was accomplished with the GALA-dendrimer-DNA complex, followed by the denrimer-DNA and then by polylysine-DNA (Table 2).

Table 2

DNA-Dendrimer Mediated Transfection

Condition	Luciferase lights (units per mg cell protein)
Dendrimer-GALA-DNA	$(9 \pm 2) \times 10^5$ ($n = 2$)
Dendrimer-DNA	$(5 \pm 2) \times 10^5$ ($n = 2$)
Polylysine-DNA	$(2.7 \pm 0.1) \times 10^5$ ($n = 2$)

Example 2

Synthesis of Reactive and Functionalized Spermidine Bis-Acridines

Spermidine bis-acridine derivatives (synthesis shown in Figure 8) intercalate into double stranded nucleic acids with affinity constants greater than 1×10^4 (pH 7.4; 0.2M NaCl) and can be used to attach a variety of targeting molecules to DNA. Carbohydrates, peptides, hormones, vitamins, cofactors, proteins or antibodies can all be used as targeting ligands.

A. Spermidine Bis-Acridine

The scheme for directing nucleic acids to certain sites of the body is based upon the intercalation of a ligand which interacts with a cell surface componentinto the double stranded DNA. A procedure for selective N^4-acylation of spermidine, using N^1,N^8-bis(t-butoxycarbonyl) spermidine as starting material, has been reported. R.J. Bergeron et al., Synthesis (1982) 689-692. We have used this procedure (Figure 8) to link the acid functionalized galactosyl derivatives 9 and 9' to the secondary amino group of N^1,N^8-Boc-protected spermidine (15) and the resulting galactosylated spermidines 17 and 17' were, after

deprotection, further alkylated with 9-phenoxyacridine by a standard chemical procedure to transform them into bis-intercalator compounds **21** and **21'**. The synthesis of the carboxylic acid functionalized galactosyl derivatives is detailed (J. Haensler et al., Biochim Biophys Acta (1988) 946:95-105) and is easily applicable to a wide range of carbohydrate ligands (M.M. Ponpipom et al., J Med Chem (1981) 24:1388-1395). The title compounds were obtained in an overall yield of 30% and the NMR and mass spectrometry data are consistent with the proposed structure.

B. Activated Spermidine Bis-Acridine

Based upon the above scheme, a versatile method for attaching peptides to a spermidine bis-acridine derivative, has been developed. N^1,N^8-bis(t-butoxycarbonyl) spermidine was N^4-acylated with N-succinimidyl-4-(p-maleimidophenyl)butyrate (SMPB) (4), deprotected and coupled to acridine rings to make a bis-intercalator bearing a maleimide group (**20**). A single compound was obtained after chromatographic purification on silicic acid in 25% overall yield. The NMR and mass spectrometry results are consistent with the assigned structure.

C. Spermidine Bis-Acridine Linked to a NLS

The NLS peptide PKKKRKV (Kaneda et al., supra) and control peptides with the same composition but a different sequence have been synthesized on an ABI automatic peptide synthesizer with an N-terminal cysteine residue. The cysteine peptide is then attached to the maleimide bearing intercalator (Figure 9) and can be anchored into double stranded nucleic acids.

D. Biodegradable Linkers

Biodegradable linkers consisting of a lys-lys peptide linkage are synthesized in the manner shown in Figure 10. In the figure, a galactose residue is placed on the unprotected amine. Alternatively, a protected peptide containing two adjacent lysine residues is synthesized by solid phase synthesis. The peptide carries membrane

permeabilization functions or targeting functions and acridine residues are added to the two ε-amino gropus on the lysines.

Example 3

Gel Retardation Assay of pC LUC4 Plasmid with Galactosylated Intercalators and Agglutinin

To demonstrate that the galactosylated bis-acridines **21** and **21'** of Example 2 (**21'** is the homolog of **21** where the galactose is separated from spermidine bis-acridine by three extra carbons) can interact with a soluble receptor while attached to DNA, we used a gel shift assay. In this assay, a galactose binding protein, *Ricinus Communis* lectin RCA_{120}, was incubated with the galactosyl-bis-acridine-DNA complex. If this protein interacts with the complex and the complex remains associated with the DNA, the DNA does not migrate into the electrophoresis gel. Each sample of the plasmid pC LUC4 (2μl; 140ng) was mixed with 13.5 pmoles of **21** or **21'** and 1μl (33.3 pmoles) of RCA_{120} was added with an excess of free galactose (1.35 nmoles) when indicated. After 30 minutes of incubation at room temperature, the samples were electrophoresed through a 0.8% agarose gel using a 0.04M Tris-Acetate buffer system (pH 7.6) and stained with ethidium bromide to visualize the DNA (Figure 11).

Intercalation of the galactosylated spermidine bis-acridines into the pC LUC4 plasmid is shown by the retardation observed for the plasmid when complexed with compounds **21'** (lane B) or **21** (lane E). Intercalation of the bis-acridine into the DNA produces a change from the supercoiled form to a relaxed circular form, which migrates slower.

The capability of the plasmid-galactose complex to bind to a soluble receptor for galactose is shown by the almost complete retardation of the complex in presence of *Ricinus Communis* lectin RCA_{120} (lane C and F). RCA_{120} is a dimer and his two binding sites selective for terminal β-D-galactosyl residues and thus too can crosslink the

plasmid-galactose complexes. The interaction of RCA_{120} with the plasmid pC LUC4 when complexed to compounds **21** or **21'** results in a formation of large aggregates which do not penetrate into the gel. This interaction appears to be much more efficient when the plasmid is complexed with **21'** than with **21**. To crosslink the plasmids, RCA_{120} has to overcome electrostatic repulsions existing between adjacent plasmids. Thus, separating the galactose from the surface of the plasmids by a spacer arm, as in case of the complexes obtained with compound **21'**, makes the binding of the lectin easier. As a result of a multivalent interaction, the plasmid aggregates formed by RCA_{120} are very stable and are not dissociated by a 100-fold excess of a competing monovalent ligand such as galactose (lanes D and G).

Example 4

Binding of Bis-acridines to Double-Stranded DNA Using Ethidium Bromide Displacement Assay

The affinity of the bis-acridines for calf thymus DNA was calculated from the displacement of ethidium bromide from double stranded nucleic acids (Nielsen, supra). Ethidium displacement was monitored by the decrease of the ethidium bromide fluorescence (ex. = 540 nm, em. = 610 nm) that occurs when it is released from DNA. The association constants of the bis-acridines relative to ethidium bromide are calculated from their IC_{50}. In this study, spermidine bis-acridine trihydrochloride (SBA·3HCl) synthesized as described (Nielsen, supra), was used as the reference compound. As a result of the loss of one of its three positive charges, a slight but significant decrease in affinity is observed when the N^4 amino group of spermidine bis-acridine is engaged in an amide bond with the targeting carbohydrate in compound **21** (Gal-BA·2HCl). However, we predict an increase in affinity when spermidine bis-acridine is linked to the highly positively charged NLS peptide PKKKRKV. G. Karup et al., Int J Peptide Protein Res (1988) 32:331-343.

conjugates synthesized to attach targeting ligands to DNA in the various examples are given in Table 3.

Table 3

Dissociation Constants of the Bis-acridines from Calf Thymus DNA (in M)

SBA·3HCl	2.4×10^{-8}
Gal-3-bA[1]	3.5×10^{-7}
Gal-6-bA[2]	7.9×10^{-7}
Gal_3Lys_2-bA[3]	5.4×10^{-6}
Maleimido-bA[4]	6.5×10^{-7}
WTcys-bA[5]	1.4×10^{-7}
SNL-bA[6]	1.4×10^{-7}

[1]Compound **21** where n=3

[2]Compound **21** where n=6

[3]Compound **26** (shown in Example 6)

[4]Compound **20**

[5]SBA linked to CGYGPKKKRKVGG

[6]SBA linked to CGYKPKVRGKGKG

The binding constants for the various bis-acridines are computed from an ethidium bromide displacement assay by using a method to determine the binding affinity of a 4-Mer for a linear lattice via noncooperative competitive binding with a 2-Mer (A. Wolfe and T. Meehan. J. Mol. Biol. 223, 1063-1087, 1992) and an intrinsic dissociation constant of 5.3×10^{-6}M for ethidium bromide.

Example 5

Ability of Bis-acridine Galactosyl Ligands to Target DNA to Cell-surface Receptors

To demonstrate the factors that control targeting ability of the bis-acridine intercalators containing a galactosyl targeting ligand, rat hepatocytes were isolated from rat liver and placed in culture at a density of 10^6 hepatocytes in 60 mm petri dishes in 3 ml of MEM medium containing 5% fetal calf serum and antibiotics. The hepatocytes are shown to have galactose receptors by

hepatocytes are shown to have galactose receptors by binding asialorosomucoid. After 18 hours at 37°C, the medium is removed and replaced with 1 ml of MEM. Then 1 ug of ^{125}I-labeled plasmid DNA complexed to either BA·3HCL, Gal-3-bA, Gal-6-bA or Gal_3-Lys_2-bA in 100 ul water was added to the culture dish. The intercalator to plasmid ratio was 500:1 or 1000:1. The cells were incubated for an additional hour at 37°C, then rinsed and the protein digested in 1 ml NaOH (1N). The cell lysate was counted for radioactivity and the protein measured. The amount of cell associated plasmid is expressed as ng of plasmid per mg of cell protein and graphed as a function of complexing agent (Figure 12). Although all three galactosyl bis acridine compounds bind to DNA (Table 2) and can interact with a soluble galactose binding protein (Example 3), only the Gal_3-Lys_2-bA was able to interact with the cell surface receptor. Thus, efficient targeting to cell surface receptors requires a longer spacer arm as provided by the Gal_3-Lys_2-bA was able to interact with the cell surface receptor (synthesis shown in Figure 13 and Example 6).

Example 6
Synthesis of a Biodegradable Bis-acridine Containing Three Targeting Ligands: Trigalactosylated Spermidine Bis-acridine

The complete synthesis of this molecule is shown in Figure 13.

Synthesis of L-Lysyl-L-Lysine bis-trifluoroacetate**(22)**:
N-ε-BOC-L-Lysine (603 mg, 2.45 mmol) and N-N-ε-bis-BOC-L-Lysine-p-nitrophenyl ester (2.28 g; 4.9 mmol)were mixed in 40 ml of N-methyl morpholine containing 640 µl of N,N diisopropylethylamine (3.7 mmol). The mixture was stirred overnight at room temperature under argon filtered to remove insoluble traces of unreacted N-ε-BOC-L-Lysine and evaporated to dryness under high vacuum. The residue was purified in a silica gel column eluted with the system $CHCl_3/CH_3OH/H_2O$ 9:1:0.1 to afford 1.22 g of pur BOC-protected Lysine dimer; yield 87%

Deprotection: To a cooled flask (dry ice) containing 700 mg (1.2 mmol) of the BOC-protected Lysine dimer were added 5 ml of TFA. The mixture was warmed to room temperature and stirred under argon. After 30 minutes stirring the trifluoroacetic acid was evaporated in vacuo. The residue was taken up in acetone and evaporated (5 times). Finally the residue was redissolved in 14 ml of water, extracted three times with 8 ml of chloroform and lyophylized to give 480 mg of the tittle compound; yield 80%.

Protected trigalactosyl lysine dimer: Synthesis of Nα-[Nα, Nε-Bis(6-(1-thio-2,3,4,6,-tetra-O-acetyl-β-D-galactopyranosyl)hexanoyl]-L-Lysyl-Nε-(6-1-thio-2,3,4,6,-tetra-O-acetyl-β-D-galactopyranosyl)hexanoyl]-L-Lysine **(23)**.

To a solution of L-Lysyl-L-Lysine bis-trifluoroacetate (400 mg; 0.8 mmol) in a 8 ml of anhydrous DMF containing 505 μl of triethylamine (3.6 mmol) was added p-nitrophenyl 6-(1-thio-2,3,4,6,-tetra-O-acetyl-β-D-galactopyranosyl)hexanoate (1.44g; 2.4 mmol). The mixture is stirred overnight under argon and evaporated to dryness. The residue was purified by chromatography on a silica gel column eluted with CHCl3/CH3OH/H2) 90:10:0.5 to give 463 mg of the title compound; yield 35%.

MS: Calculated for $C_{73}H_{112}N_4O_{32}S_3$ m/z= 1652, found m/z = 1653.6 (M+H)+, m/z= 1677.6 (M+Na)+, m/z = 1693.6 (M+K)+

Reaction with selectively blocked spermidine:

N4-[Nα-[Nα,Nε-Bis(6-(1-thio-2,3,4,6,-tetra-O-acetyl-β-D-galactopyranosyl)hexanoyl]-L-Lysyl-Nε-(6-1-thio-2,3,4,6,-tetra-O-acetyl-β-D-galactopyranosyl)hexanoyl]-L-Lysyl]-N1,N8-bis-BOC-spermidine. **(24)**.

Compound **23** (132 mg; 80 μmol) was activated by esterification with N-Hydroxysuccinimide (11 mg, 96 μmol) in the presence of DCC (20 mg; 97 μmol) in 5 ml of anhydrous methylene chloride. After 4 h stirring at room temperature under argon, the urea precipitate was removed by filtration and the filtrate was evaporated in vacuo.

The dry residue was redissolved in 3 ml of acetonitrile and added dropwise to a solution of N1,N8-bis(t-butoxycarbonyl spermidine) Hydrochloride (30 mg; 80 μmol) in 3 ml of acetonitrile containing 14 μl of triethylamine (100 μl). The mixture was further stirred for 48 h at room temperature under argon, evaporated in vacuo to a residue which was purified in a silica gel column eluted with $CHCl_3/CH_3OH/H_2O$ 90:10:1 to afford 71 mg of the title compound; yield 45%.

Deprotection: Synthesis of N4-[Nα-[Nα,Nε-Bis(6-(1-thio-β-D-galactopyranosyl)hexanoyl]-L-Lysyl-Nε-(6-(1-thio-β-D-galactopyranosyl)hexanoyl]-L-Lysyl] spermidine. **(25)** Compound **24** (71 mg; 36 μmol) was deprotected as described previously for compound 11. The BOC protecting groups were removed from the spermidine linker by treating with 5 ml of TFA for 30 min and the acetyl protecting groups were removed from the galactosyl headgroups by treating overnight with a mixture of $CH_3OH/NEt_3/H_2O$ 5:4:1. The bis-trifluoroacetate salt of the spermidine derivative was converted to the free amine by passing a water solution of the salt through a small BIO-RAD AG 1X2 (OH^-) column. The fractions positive for carbohydrates and for amines were pooled together and lyophylized.

36 mg of compound **25** was obtained; yield 79%.

Acridine attachment: Synthesis of N4-[Nα-[Nα,Nε-Bis(6-(1-thio-β-D-galactopyranosyl)hexanoyl]-L-Lysyl-Nε-(6-(1-thio-β-D-galctopyranosyl)hexanoyl]-L-Lysyl]-N1,N8-bis-acridine spermidine. **(26)** ("Gal_3-Lys_2-bA")

Compound **25** (36 mg; 28.5 μmol) and 18 mg of 9-phenoxyacridine were dissolved in 3g of phenol at 80°C and the solution was further stirred for 2 h at 80°C under argon. The mixture was then cooled to about 40°C and poured into 15 ml of ether to precipitate the aminoacridines. The yellow precipitate was collected by filtration on a filter paper and redissolved in 4 ml of a butanol/methanol mixture 3:1. This solution was then condensed by evaporation to about 1 ml and the bis-acridine derivative was isolated by chromatography on a silica gel

6:2:1:2.

34 mg of the title compound are obtained; yield 21%.

MS: Calculated for $C_{81}H_{117}N_9O_{20}S_3$ m/z = 1631, found m/z = 1632.8 (M+H)+, m/z = 1654.8 (M+Na)+.

Example 7

Transfection Assay Using Nuclear Localization Sequences

5 μl of TE containing a trace amount of a 5Kb radioiodinated plasmid (CMV-βGal) and 50 μl of water containing 8 nmoles of the nuclear localization peptide-bis-acridine conjugate of Example 2-C were added to 80 μg of pCLUC4 (123 neq. bp) in solution in 45μl of TE buffer (pH 8). The ratio of plasmid to peptide conjugate was 1:300. After 1 hour standing at room temperature 100 μl of Tris-Cl buffer (pH 9) was added to the complex and the resulting solution was mixed with 12 μmoles of lipids (DOPE/CHEMS 2:1, molar ratio) dissolved in 600 μl of ether for the preparation of pH-sensitive liposomes.

The vesicles containing the DNA-peptide complexes were separated from nonencapsulated material by floating the liposomes through a ficoll gradient. Encapsulation efficiency (20% ± 4%) was determinated by dynamic light scattering (Coulter N4, Coultronics).

Cells were transfected with 4 μl of liposome-encapsulated plasmid (100 μl of the liposome solution) for 5 hours at 37°C and luciferase activity was counted after 48 hours in a bioluminometer. Table 4 shows the measured light units/mg cell protein as a function of the liposomal content. The values are the averages of three determinations.

Table 4

Liposomal Content

plasmid alone complex	plasmid-WTcys-bA complex	plasmid-SNL-bA
(0.32 ± 0.02) 10^6	(0.82 ± 0.36) 10^6	(1.36 ± 0.28) 10^6

Positive control: Lipofectin™ = (1.4 ± 0.2) 10^8.

If we admit the pH-sensitive liposomes deliver their content into the cytoplasm of the host cell, the naked plasmid must be able to penetrate the nucleus.

If we exclude that the peptide-bis-acridines conjugates protect DNA from degradation, the observed transfection enhancement must be a result of an increased nuclear entry. The 4-5 fold increase of transfection agrees with published results (Kaneda et al., supra) using proteins that bind to DNA to enhance DNA entry into the nucleus. Both the SNL peptide and WTcys peptide increase expression and are a convenient technique to target DNA into the nucleus.

Example 8

Synthesis of Cationic Bile Salts

A. Preparation of the α-cholic Acid Amide of α-benyzlester ε-TBOC-amino-lysine

The synthesis is based upon that of S. Bergstrom et al., Acta Chem Scand (1953) 7:1126. 204 mg (0.500 millimoles) of cholic acid was weighed into a screw-capped test-tube, and 2.5 ml dioxane and 70 microliters (0.500 millimoles) of triethylamine was added to the tube. The mixture was cooled in an ice bath until the solution solidified (at about 12°C). 65 microliters (0.500 millimoles) of isobutyl chloroformate were added, the reaction tube was agitated and returned to the ice bath. The tube was alternatively removed and replaced in the bath

to keep the temperature at the point of incipient solidification for 30 minutes.

α-benyzlester N-ε-TBOC lysine (0.500 millimoles) and 70 microliters (0.500 millimoles) of triethylamine were suspended in 0.6 ml of water. The mixture was cooled in the ice bath, added to the dioxane reaction mixture, and the container rinsed into the reaction mixture with another 0.5 ml of ice water. The tube stood in the ice bath for 1/2 hour and then permitted to warm to room temperature.

Most of the organic solvent was evaporated beneath a stream of argon gas, and the residue was brought up to 3 ml with water. 5% aqueous sodium carbonate was added in a dropwise fashion until the pH reached 9. The mixture was extracted with three successive 3 ml portions of ethyl ether, and the aqueous phase saved.

To the aqueous residue, 0.5 N hydrochloric acid was added until the pH fell to 4. The mixture was extracted with three successive 3 ml portions of ethyl ether, and the aqueous phase saved.

The pH of the aqueous residue was readjusted to 4 with 0.5 N hydrochloric acid and extracted into five successive 3 ml portions of ethylacetate. These ethylacetate extracts were combined and evaporated to dryness under vacuum to obtain 284 mg of colorless powder melting. The tBoc protecting group for the ε-amine was removed by standard methods to yield the positively charged lysine derivative of cholic acid. In a similar fashion other positively charged derivatives of cholic acid can be prepared.

B. Preparation of cholic acid amide of tris (2-aminoethyl)amine

When multiple amine groups are available for coupling to the activated cholic acid the amine is added in a 6 fold excess over the activated bile salt prepared as described in Example 8-A. The synthesis is based upon that of Bergstrom et al., supra. Weigh 204 mg (0.500 millimoles) of cholic acid into a screw-capped test tube. Add 2.5 ml dioxane and 70 microliters (0.500 millimoles) of triethyl

amine. Cool in an ice bath until the solution commences to solidify (at about 12'C). Add 65 microliters (0.500 millimoles) of isobutyl chloroformate and after agitating, return the reaction tube to the ice bath. By alternately removing from the ice bath and replacing in the bath, keep the temperature at the point of incipient solidification for 30 minutes.

Add (3.00 millimoles) of tris(2-aminoethyl)amine and 70 microliters (0.500 millimoles) of triethylamine in 0.6 ml of water. Cool in the ice bath, add to the dioxane reaction mixture, and rinse the container into the reaction mixture with another 0.5 ml of ice water. Let stand in the ice bath for 1/2 hour and then permit to warm to room temperature.

Evaporate most of the organic solvent beneath a stream of argon gas. Make the residue back up to 3 ml with water. Add 5% aqueous sodium carbonate dropwise until the pH reaches 7. Extract with three successive 3 ml portions of ethyl ether, saving the aqueous phase.

To the aqueous residue add 0.5 N hydrochloric acid until the pH falls to 4. Extract with three successive 3 ml portions of ethylether, saving the aqueous phase.

Readjust the pH of the aqueous residue to 4 with 0.5 N hydrochloric acid and extract into five successive 3 ml portions of ethylacetate. Combine these ethyl acetate extracts and evaporate to dryness under vacuum to obtain the cholic acid amide of tris(aminoethyl)amine.

Example 9

Synthesis of Polyethyleneglycol-Bis Acridine

The synthesis of PEG-coupled bis-acridine spermidine follows standard chemistry and involves the following steps:

$R'O-(CH2-CH2-O)_n-H$ -----Activating agent--->

$R'O(CH2-CH2-O)_n-R^*$ -----bis-acridine------->

$R'O-(CH2-CH2-O)_n-R$-Bis acridine where R' = H or CH_3 and R^* = activating group and n = 10-250, preferably 20-60.

There are many methods for preparing activated

monomethoxy PEG molecules or activated PEG molecules. A preferred method has been described by D. Larwood and F. Szoka, J Labelled Comp & Radiopharm (1984) 21:603-614. Polyethylene glycol 1900 carbonyl-imidazole methyl ether was prepared by taking 530 mg (0.28 mmol) dry PEG 1900 monomethyl ether in 2 ml dry methylene chloride and adding 78 mg (0.46 mmol) carbonyldiimidazole and 10 mg (0.11 mmole) imidazole (sodium salt). After stirring overnight, 6 ml dry methylene chloride was added and the mixture extracted with 3.75 ml water, then dried with anhydrous sodium sulfate. After filtration, solvent was removed, with quantitative yield. Alternatively, the solvent was removed, and the resulting oil recrystallized from chloroform/diethyl ether at -20°C. The resulting imidazole carbamate white crystals were filtered through a chilled funnel, rinsed with a small amount of diethyl ether, and used immediately.

The imidazole carbamate (0.1mM) is added to 0.125 mM of N,N'-bis-(9-acridinyl)-4-aza-1,8-diaminooctane ("bis acridine-spermidine", prepared as described by P. Nielsen, Eur. J. Biochem. 122:283-289, 1992), dissolved in phenol and the reaction run at 80°C under argon for 2 hr. The mixture is taken to dryness and the yellow product washed with cold ethanol and then diethyl ether. The PEG is coupled via a carbamate linkage to the secondary amine of the bis-acridine spermidine to form the monomethoxy PEG-bis acridine spermidine and is soluble in water.

In a similar fashion the non-blocked PEG (molecular weight 6000), is activated as above to form the bis-imidazole carbamate PEG. The bis-imidazole carabamate PEG is reacted with a 2.5 fold excess of bis-acridine spermidine to form the bis(bis-acridine-spermidine)-PEG 6000.

Various types of activators for PEG and monomethoxy PEG have been described in U.S. patent 5,013,556 to Woodle et al. These methods can be used to generate reactive PEGs that can be attached to the bis-acridine molecule via a variety of chemistries. For instance a sulfhydryl

containing monomethoxy-PEG can be attached to the maleimide-containing bis-acridine of Example 2-B.

Example 10

DNA-Masking with PEG-bis-acridine

PEG molecules can be used to mask the surface of the DNA and permit the DNA to circulate for a longer period. Radio-iodinated plasmid DNA is mixed with monomethoxy-PEG-1900-bis-acridine spermidine as synthesized in Example 9 at a 20 bp DNA-to-1 PEG molecule ratio, for 30 minutes at room temperature. An aliquot of the complex, 5 ug DNA in 0.2 ml PBS, is injected via the tail vein into each of a group of 12 mice. Mice are sacrificed at various periods after injection. The blood and other organs are removed and the radioactivity associated with each organ is determined. DNA which has not been complexed to the monomethoxy-PEG-bis-acridine-spermidine is injected into a second group of mice, designated the control mice. Evidence showing that after 10 minutes, 15% of the radioactive plasmid DNA remains in the blood in the control mice, whereas in the monomethoxy PEG-bis acridine spermidine DNA group significantly greater levels of the radiolabeled plasmid-PEG complex remain in circulation indicates a pronounced masking effect of the DNA molecule by the PEG bis acridine spermidine.

Example 11

Synthesis of a Lecithin Acyl Amine Masking Reagent

The synthesis of polynucleotide masking lipids is accomplished by standard chemistry such as that described in C. Pidgeon et al., Anal Biochem (1989) 176:36-47.

1. 1,12-dodecanedicarboxylic acid + DCC

--- THF, 25°C ------>

dodecanedicarboxylic acid anhydride (cyclic anhydride)

2. monoacyl lysolecithin + cyclic anhydride

----- $CHCl_3$, DMAP, 25°C 48 hr ------>

Lecithin - COOH

3. Lecithin-COOH + carbonyldiimidazole

--- $CHCl_3$, 25°C, 2 hr --->

Lecithin imidazolide

4. Lecithin imidazolide + amine reactant

--- $CHCl_3$, 25°C, 24 hr --->

cationic lecithin

The final reaction of the amine reactant with the lecithin imidazolide is undertaken immediately after formation of the lecithin imidazolide. The lecithin imidazolide (0.1 mM) is added in to a solution of the amine (0.7 mM) in chloroform. Suitable amines for this coupling are listed in the specification.

After two hours at room temperature the reaction mixture is added to a two fold volume of water/methanol and the pH is adjusted to 10. The lecithin linked amine is extracted into the organic phase. The organic phase is then washed with 0.1 M sodium chloride and the organic phase taken to dryness. The resulting acyl amine lecithin is used to mask the surface of the polynucleotide. Various lysolecithin molecules can be used to prepare the lecithin-COOH, including dodecyl, myristoyl, palmitoyl, oleyl or phytanyl, or stearyl. Other headgroups such as ethanolamine or phosphatidic acid can be substituted for lecithin if they are suitably protected in the activation steps and deprotected at the end of the reaction.

Example 12

DNA-Masking with Lecithin Acyl Amine

The lecithin acyl amine of Example 11 can be added to DNA from an ethanol solution at a ratio of 1 positive charge to each phosphate group on the DNA. The molecule

can be used to mask the surface of the DNA and permit the DNA to circulate for a longer period. An aliquot of the complex, 5 ug DNA in 0.2 ml PBS, is injected via the tail vein into each of a group of 12 mice. Mice are sacrificed at various periods after injection. The blood and other organs are removed and the radioactivity associated with each organ is determined. DNA which has not been complexed to the lecithin acyl amine is injected into a second group of mice, designated the control mice. Evidence showing that after 10 minutes, 15% of the radioactive plasmid DNA remains in the blood in the control mice, whereas in the monomethoxy PEG-bis acridine spermidine DNA group significantly greater levels of the radiolabeled plasmid-PEG complex remain in circulation indicates a pronounced masking effect of the DNA molecule by the lecithin acyl amine.

CLAIMS

We claim:

1. A composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell, said composition comprising the polynucleotide associated with a membrane-permeabilizing component capable of transporting the polynucleotide across the cytoplasmic membrane of said eukaryotic cell.

2. The composition of claim 1 wherein said membrane-permeabilizing component is an amphipathic cationic peptide.

3. The composition of claim 2 wherein said peptide is a cyclic peptide.

4. The composition of claim 3 wherein said cyclic peptide is selected from the group consisting of gramicidin S and tyrocidines.

5. The composition of claim 4 wherein said cyclic peptide is gramicidin S.

6. The composition of claim 1 further comprising a phospholipid.

7. The composition of claim 6 wherein said phospholipid is a phosphatidylethanolamine.

8. The composition of claim 7 wherein said phosphatidyl ethanolamine is dioleoyl phosphatidyl-ethanolamine.

9. The composition of claim 6 wherein said phospholipid is in the form of liposomes.

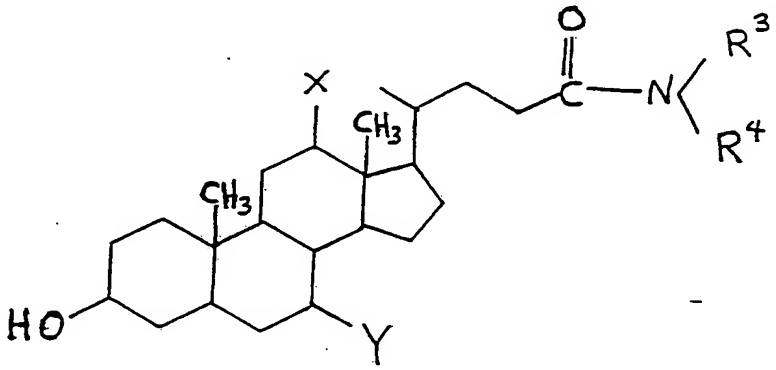
10. The composition of claim 1 further comprising a polycation.

11. The composition of claim 10 wherein said polycation is a polyamine.

12. The composition of claim 11 wherein said polyamine is selected from the group consisting of spermine, spermidine, 3,3'-diamino-bispropylamine, iminobis(N,N)-dimethylpropylamine, iminobis(3-aminopropyl)-1,3-propanediamine, and cationic dendrimers.

13. The composition of claim 6 further comprising a polycation.

14. The composition of claim 1 wherein said membrane-permeabilizing component is a cationic bile salt having the formula



wherein X and Y are independently H or OH;

R^3 is hydrogen, alkyl or alkylamine having from 1 to 10 carbon atoms; and

R^4 is a positively charged linear or branched alkyl or alkylamine having from 1 to 30 carbon atoms, wherein one or more of the carbon atoms may be substituted with NR', wherein R' is hydrogen, alkyl or alkylamine having from 1 to 10 carbons.

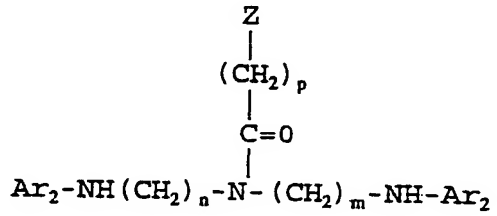
15. A composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell comprising the polynucleotide associated with a cell recognition component capable of recognizing said eukaryotic cell.

16. The composition of claim 15 wherein said cell recognition component comprises a ligand for a cell-surface receptor on said eukaryotic cell, said ligand being coupled to a DNA-associating moiety.

17. The composition of claim 16 wherein said DNA-associating moiety is an intercalating agent.

18. The composition of claim 16 wherein said DNA-associating moiety is a major- or minor-groove binder.

19. The composition of claim 17 wherein said intercalating agent has the formula



wherein

Z is a bond;

each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20; and

Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof.

20. The composition of claim 19 wherein said intercalating agent is coupled to a plurality of ligands.

21. The trigalactosylated spermidine bis-acridine compound (26) as shown in Figure 13.

22. The composition of claim 15 further comprising a membrane-permeabilization component.

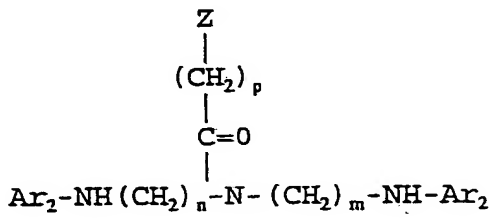
23. A composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell comprising the polynucleotide associated with a subcellular-localization component capable of delivering the polynucleotide from the cytoplasm of said eukaryotic cell to a subcellular component of said eukaryotic cell.

24. The composition of claim 23 wherein said subcellular component is the nucleus, and said subcellular-localization component is a nuclear localization component.

25. The composition of claim 24 wherein said nuclear-localization component comprises a nuclear localization sequence coupled to a DNA-associating moiety.

26. The composition of claim 25 wherein said DNA-associating moiety is an intercalating agent.

27. The composition of claim 26 wherein said intercalating agent has the formula



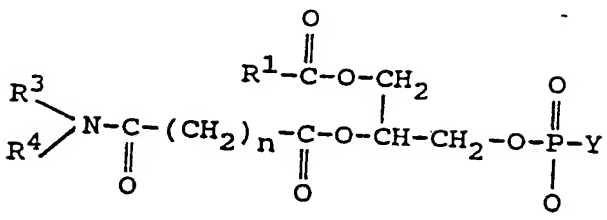
wherein

Z is a bond;

each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20; and

Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof.

28. A DNA-masking component having the formula



wherein n is an integer from 6 to 24;

Y is selected from the group consisting of hydroxy, ethanolamine, choline, glycerol, serine and inositol;

R^1 is alkyl or alkenyl having from 6 to 24 carbon atoms;

R^3 is hydrogen, alkyl or alkylamine having from 1 to 10 carbon atoms; and

R^4 is a positively charged linear or branched alkyl or alkylamine having from 1 to 30 carbon atoms, wherein one or more of the carbon atoms may be substituted with NR', wherein R' is hydrogen, alkyl or alkylamine having from 1 to 10 carbons.

29. A composition for presenting a polynucleotide to a subcellular component of a eukaryotic cell comprising

(a) the polynucleotide;

(b) a cell recognition component capable of recognizing said eukaryotic cell;

(c) a membrane-permeabilizing component capable of transporting the polynucleotide across the cytoplasmic membrane of said eukaryotic cell; and

(d) a nuclear-localization component capable of delivering the polynucleotide from the cytoplasm of said eukaryotic cell to a subcellular component of said eukaryotic cell.

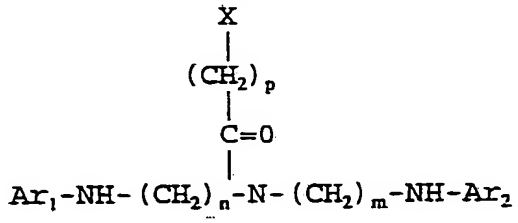
30. The composition of claim 29 further comprising:

(e) a masking component capable of increasing the circulatory half-life of the polynucleotide.

31. The composition of claim 30 wherein said masking component is polyethylene glycol (PEG) covalently linked to a DNA-associating moiety.

32. The composition of claim 30 wherein one or more of the cell-recognition component, the membrane-permeabilizing component, the subcellular-localization component and the masking component are capable of dissociation from the polynucleotide.

33. An intercalating compound having the formula



wherein

each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20;

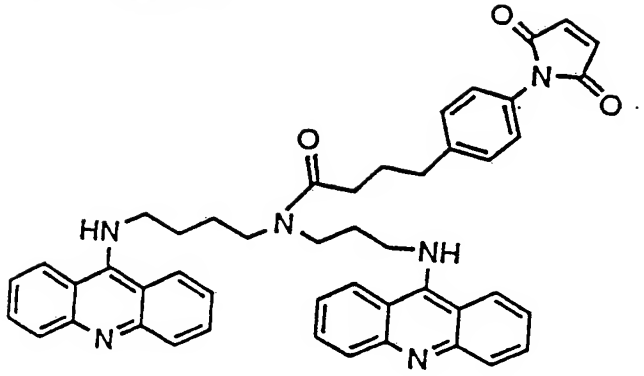
Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof; and

X is a reactive group selected from the group consisting of N-hydroxysuccinimide, maleimide, maleimidophenyl, pyridyl disulfide, hydrazide, and phenylglyoxal.

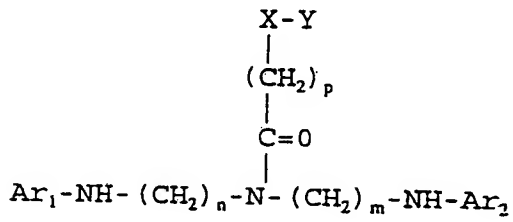
34. The intercalating compound of claim 33 wherein Ar_1 and Ar_2 are acridines.

35. The intercalating compound of claim 34 wherein X is maleimidophenyl.

36. The compound



37. A compound having the formula



wherein

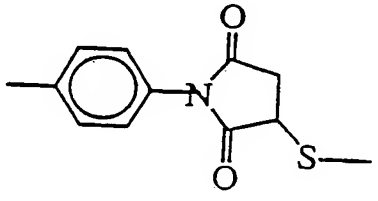
each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20;

Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof;

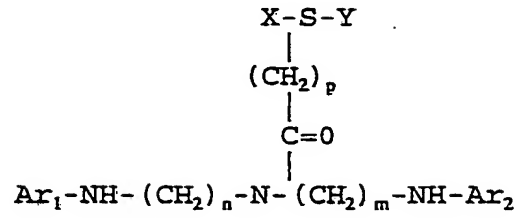
X is a reactive group selected from the group consisting of N-hydroxysuccinimide, maleimide, maleimidophenyl, pyridyl disulfide, hydrazide, and phenylglyoxal; and

Y is selected from the group consisting of cell surface receptor ligands, nuclear localization sequences, and membrane permeabilizing components.

38. The compound of claim 37 wherein X is



39. A method for making the compound having the formula



wherein

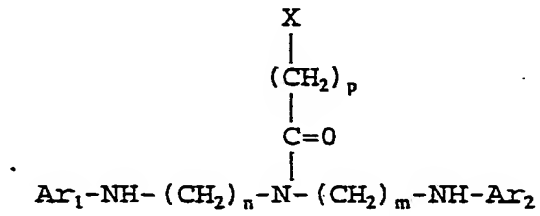
each of n and m is independently an integer of 1 to 20, and p is an integer of 0 to 20;

X is maleimidophenyl;

Ar_1 and Ar_2 are independently selected from the group consisting of ethidium bromide, acridine, mitoxantrone, oxazolopyridocarbazole, ellipticine and N-methyl-2,7-diazapyrenium, and derivatives thereof; and

Y is selected from the group consisting of cell surface receptor ligands, subcellular localization sequences, and membrane permeabilizing components,

said method comprising reacting a component having the formula



with a free sulfhydryl group on compound Y.

40. A method for introducing polynucleotides into cells in vitro comprising contacting said cells with the composition of claim 29.

41. A method for introducing polynucleotides into cells in vivo comprising contacting said cells with the composition of claim 29.

42. A method for introducing polynucleotides into plant cells comprising contacting said cells with the composition of claim 29.

43. A method for introducing polynucleotides into mammalian cells comprising contacting said cells with the composition of claim 29.

44. A method for introducing polynucleotides into the lung of a mammal comprising administering an aerosol composition including the composition of claim 29 to the lung of the mammal.

45. A method for gene therapy in a human patient comprising administering to said patient a composition of claim 29.

46. The composition of claim 16 wherein said DNA-associating moiety is a linker strand.

47. The composition of claim 16 wherein said DNA-associating moiety is a dendrimer polycation.

48. The composition of claim 1 wherein said membrane-permeabilizing component is an amphipathic peptide such as GALA.

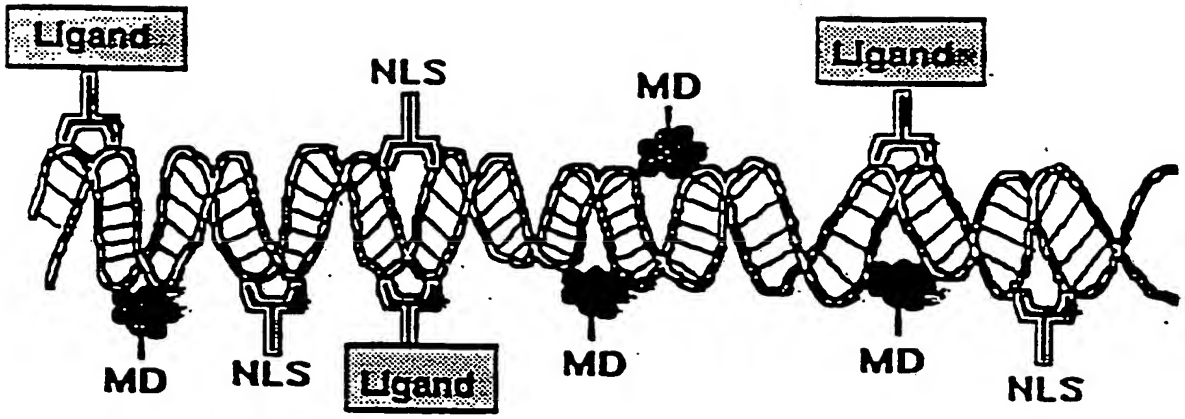


Fig. 1

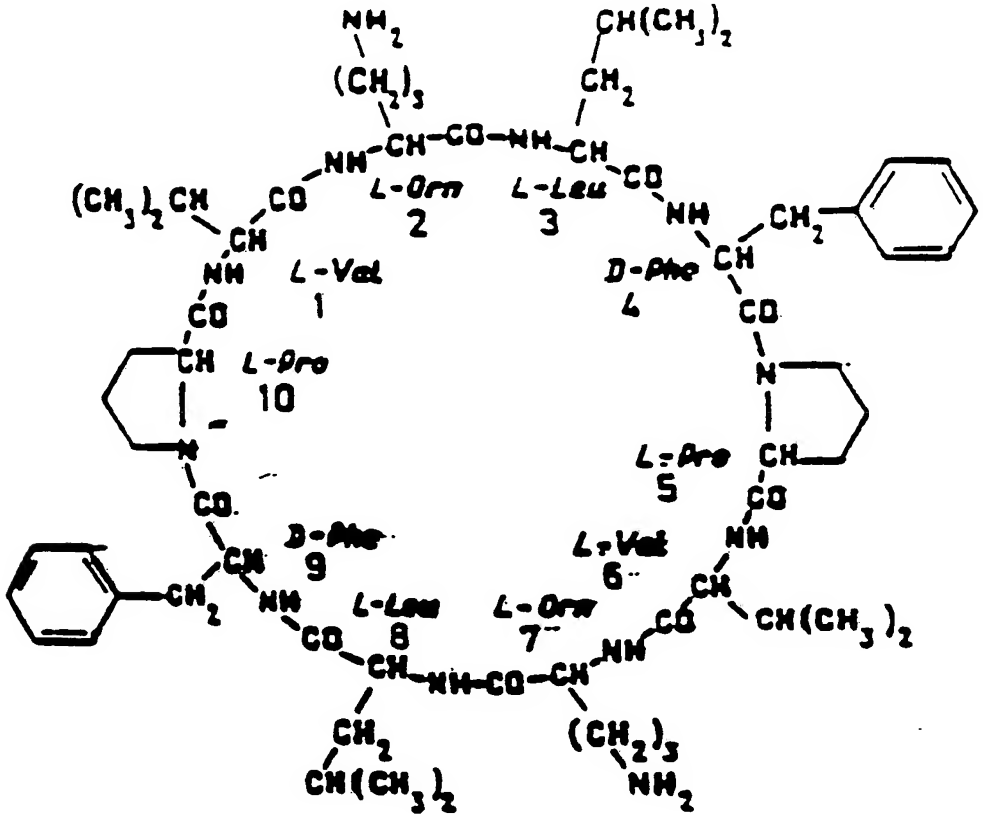

L-Val
1
L-Orn
2
L-Leu
3
D-Phe
4
L-Pro
5
L-Val
6
L-Orn
7
L-Leu
8
D-Phe
9
L-Pro
10

Fig. 2

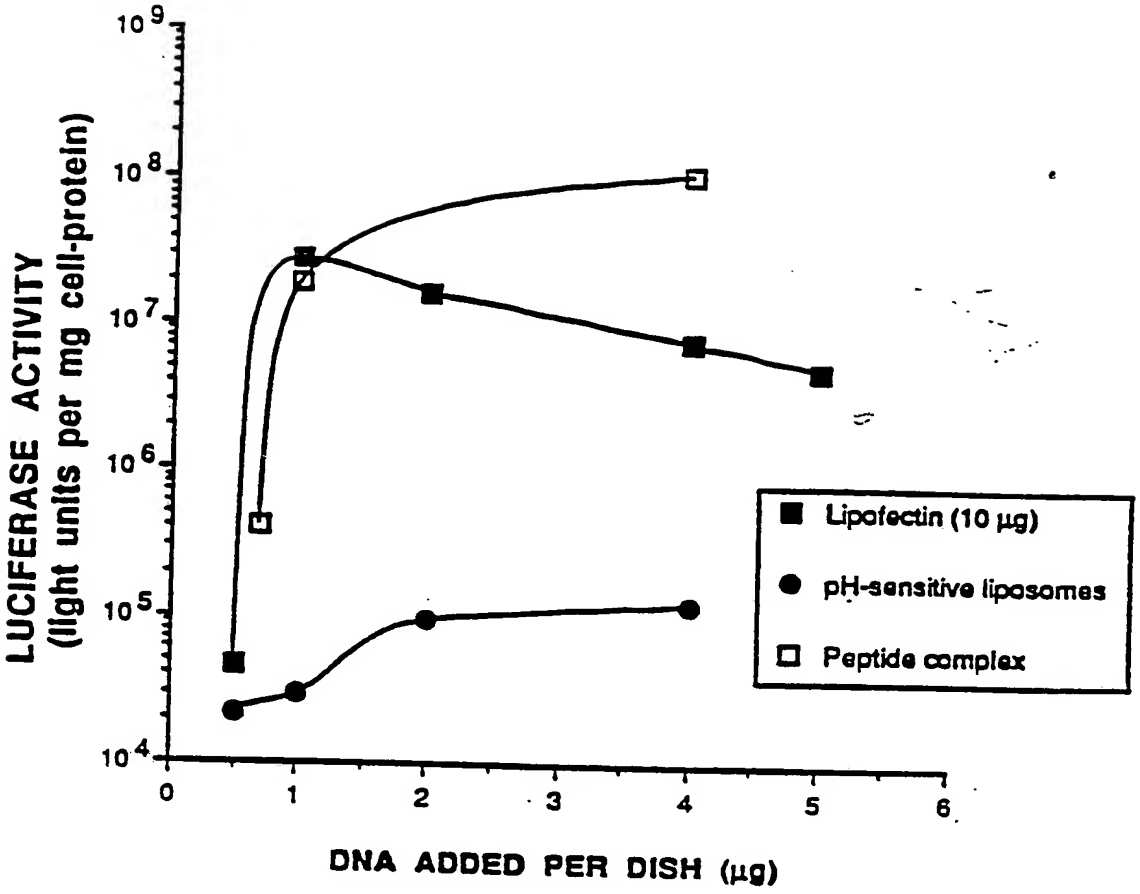

LUCIFERASE ACTIVITY
(light units per mg cell-protein)
10 9
10 8
10 7
10 6
10 5
10 4
0
1
2
3
4
5
6
DNA ADDED PER DISH (µg)
Lipofectin (10 µg)
pH-sensitive liposomes
Peptide complex

Fig. 3

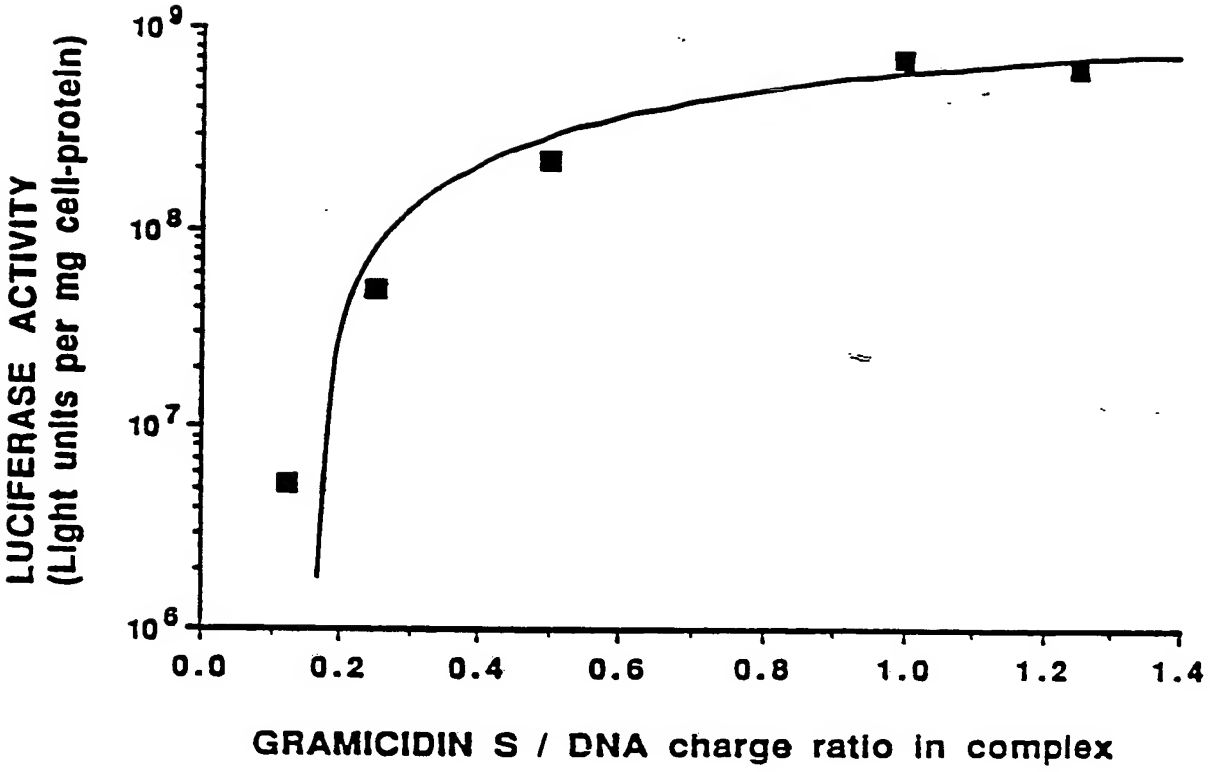

LUCIFERASE ACTIVITY
(Light units per mg cell-protein)
10^9
10^8
10^7
10^6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
GRAMICIDIN S / DNA charge ratio in complex

Fig. 4

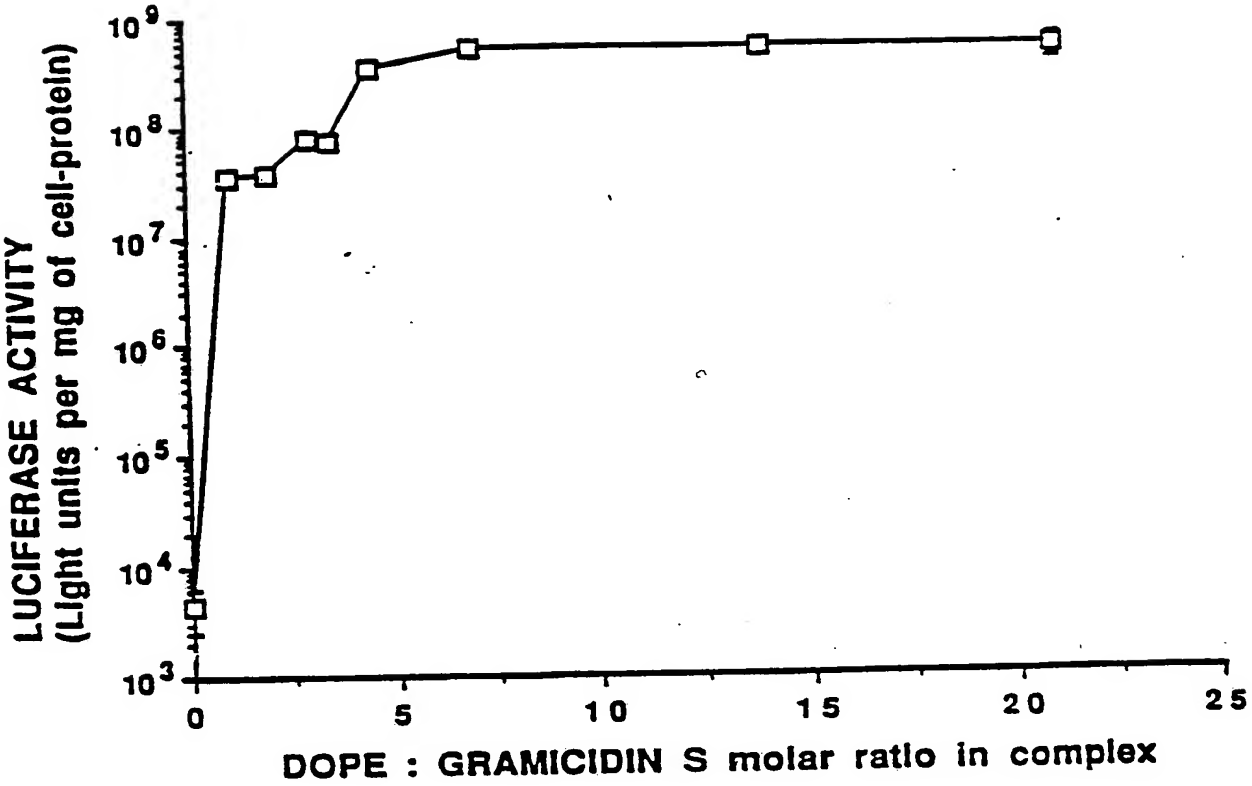

LUCIFERASE ACTIVITY
(Light units per mg of cell-protein)
10^9
10^8
10^7
10^6
10^5
10^4
10^3
0
5
10
15
20
25
DOPE : GRAMICIDIN S molar ratio in complex

Fig. 5

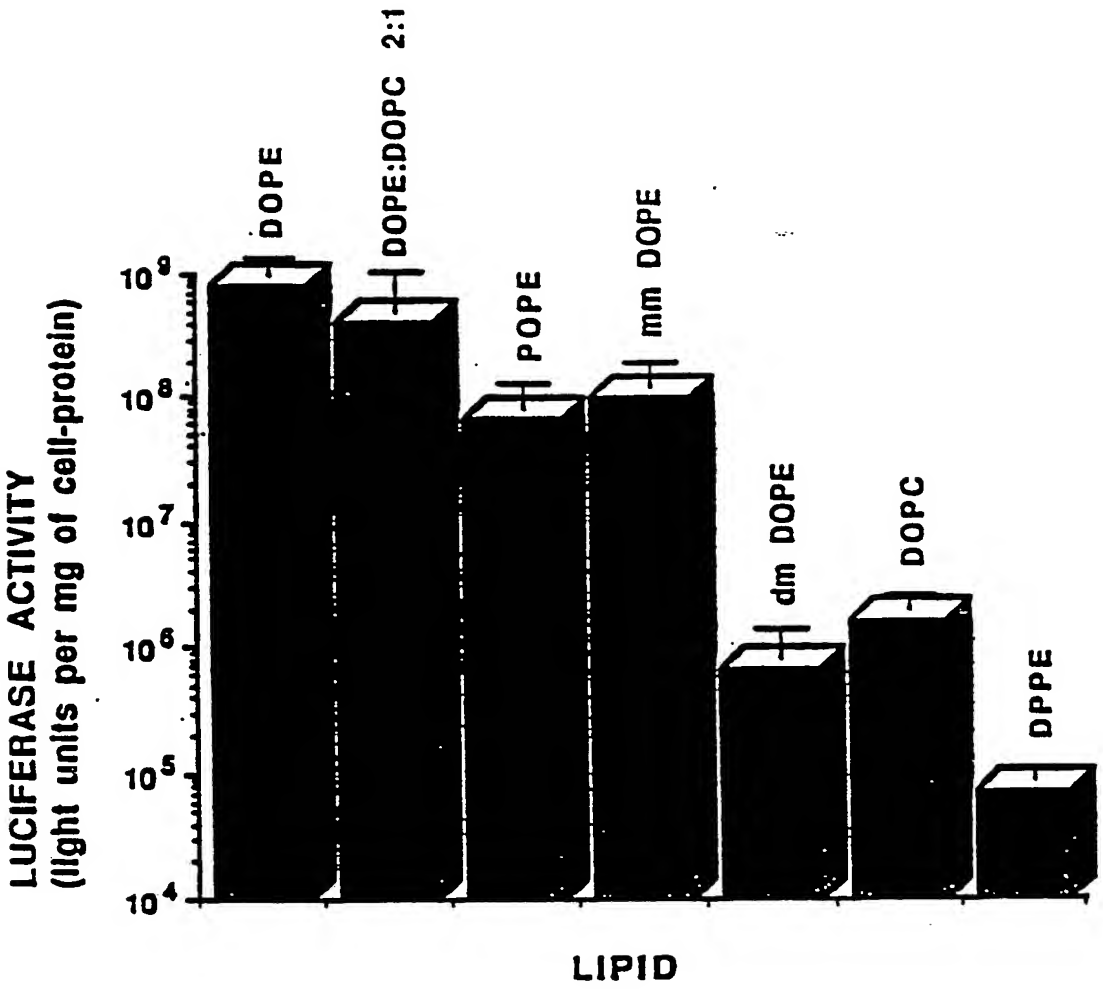

LUCIFERASE ACTIVITY
(light units per mg of cell-protein)
10^9
10^8
10^7
10^6
10^5
10^4
DOPE
DOPE:DOPC 2:1
POPE
mm DOPE
dm DOPE
DOPC
DPPE
LIPID

Fig. 6

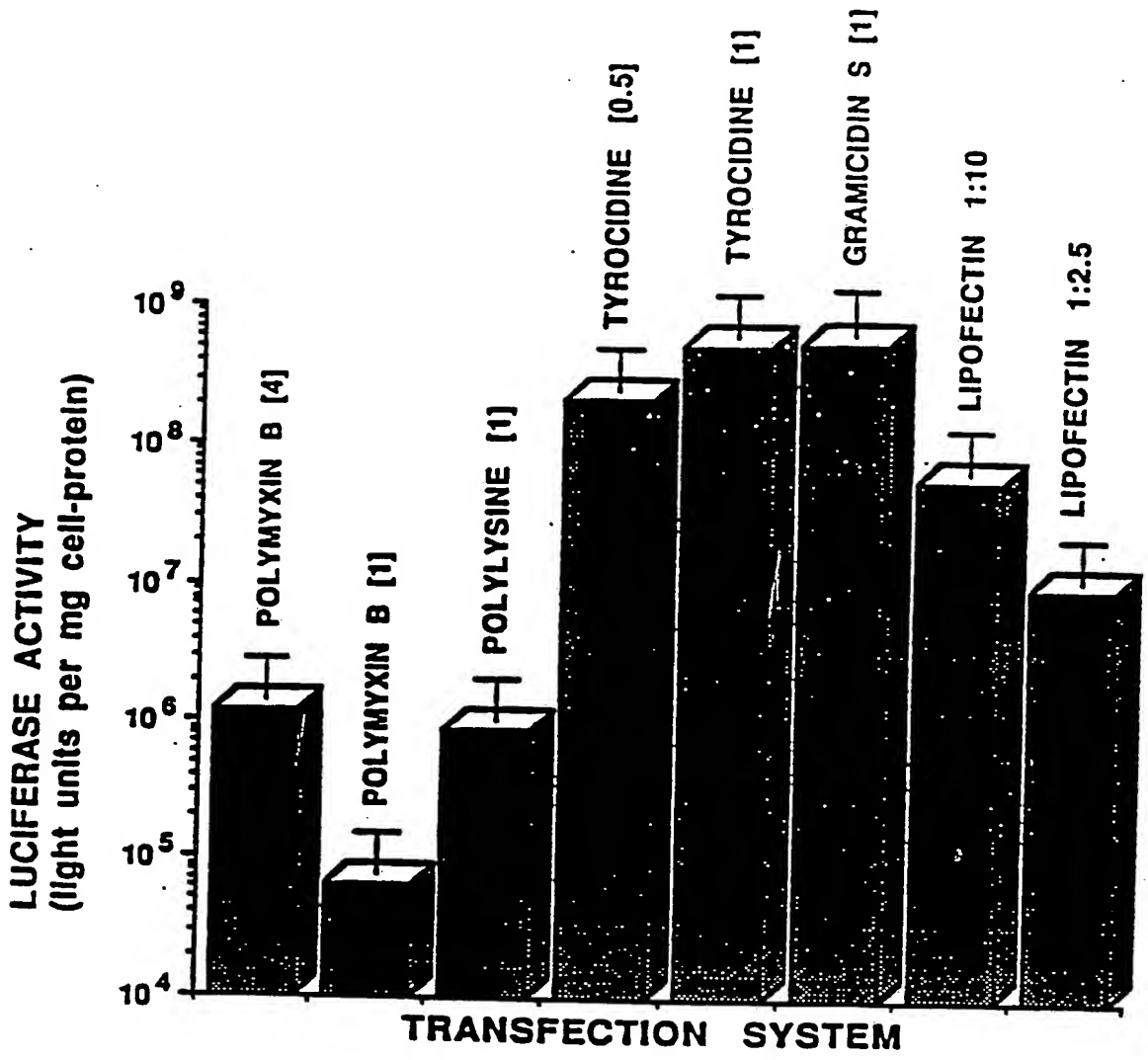

LUCIFERASE ACTIVITY
(light units per mg cell-protein)
10^9
10^8
10^7
10^6
10^5
10^4
POLYMYXIN B [4]
POLYMYXIN B [1]
POLYLYSINE [1]
TYROCIDINE [0.5]
TYROCIDINE [1]
GRAMICIDIN S [1]
LIPOFECTIN 1:10
LIPOFECTIN 1:2.5
TRANSFECTION SYSTEM

Fig. 7

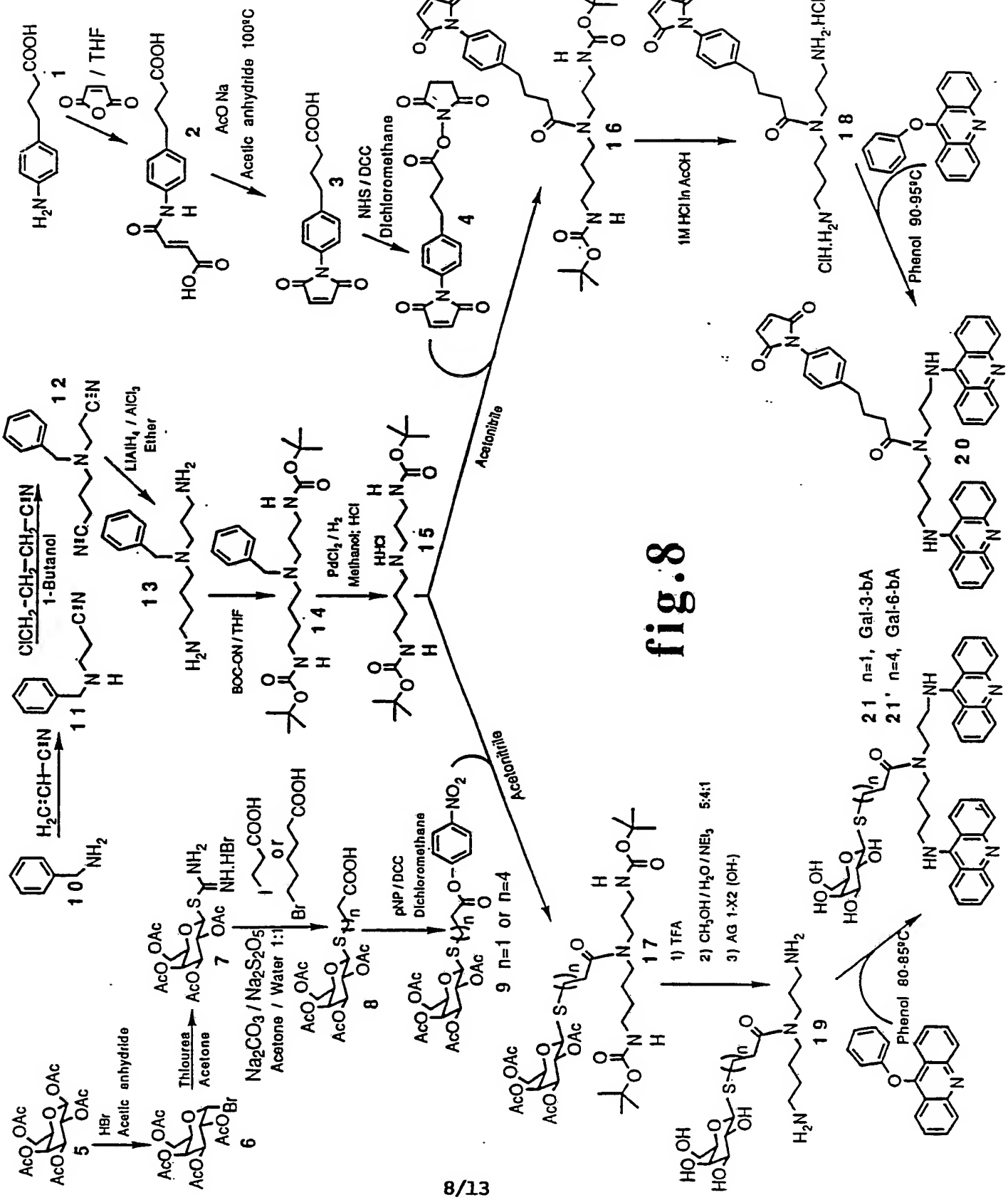


fig.8

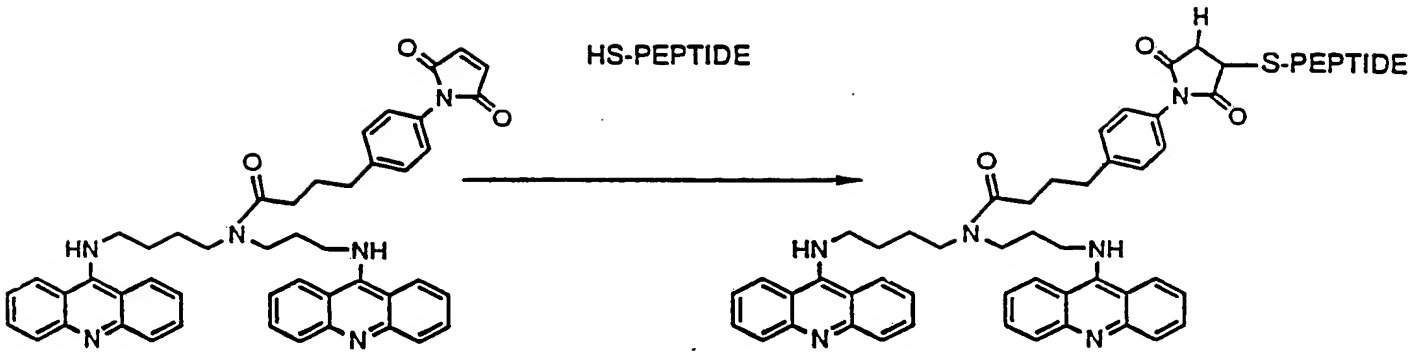


fig.9

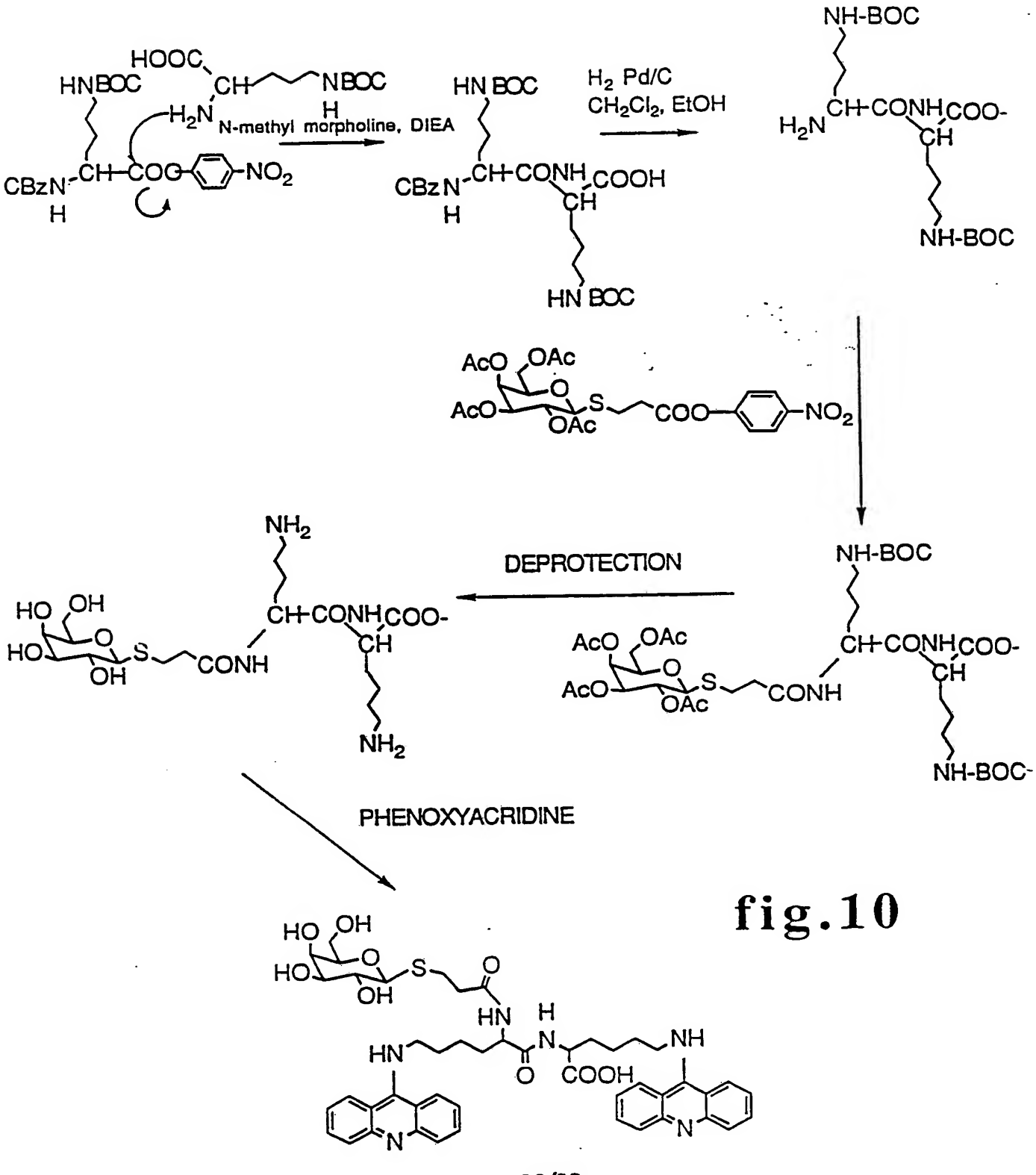

HOOC
HNBOC
CH
NBOC
H
H_2N
N-methyl morpholine, DIEA
CBzN
H
CH
COO
NO_2
HN BOC
H_2 Pd/C
CH_2Cl_2, EtOH
NH-BOC
H_2N
CH-CONH COO-
CH
NH-BOC
CBzN
H
CH-CONH COOH
CH
HN BOC
AcO
OAc
O
AcO
S
OAc
COO
NO_2
DEPROTECTION
NH_2
HO
OH
O
HO
S
OH
CONH
CH-CONH COO-
CH
NH_2
NH-BOC
AcO
OAc
O
AcO
S
OAc
CONH
CH-CONH COO-
CH
NH-BOC
PHENOXYACRIDINE
HO
OH
O
HO
S
OH
O
HN
H
N
HN
NH
O
COOH
N
N

fig.10

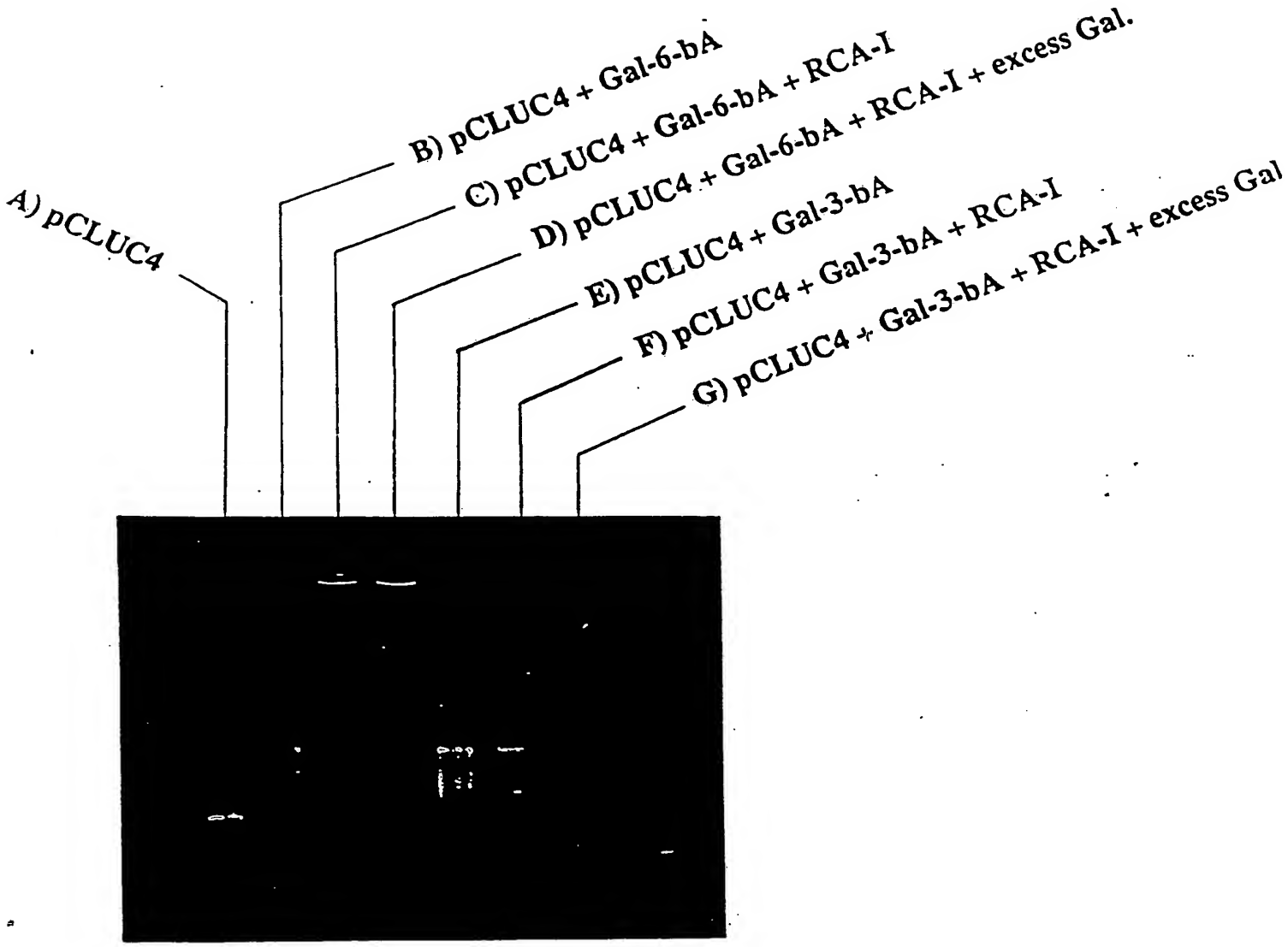

A) pCLUC4
B) pCLUC4 + Gal-6-bA
C) pCLUC4 + Gal-6-bA + RCA-I
D) pCLUC4 + Gal-6-bA + RCA-I + excess Gal.
E) pCLUC4 + Gal-3-bA
F) pCLUC4 + Gal-3-bA + RCA-I
G) pCLUC4 + Gal-3-bA + RCA-I + excess Gal

fig.11

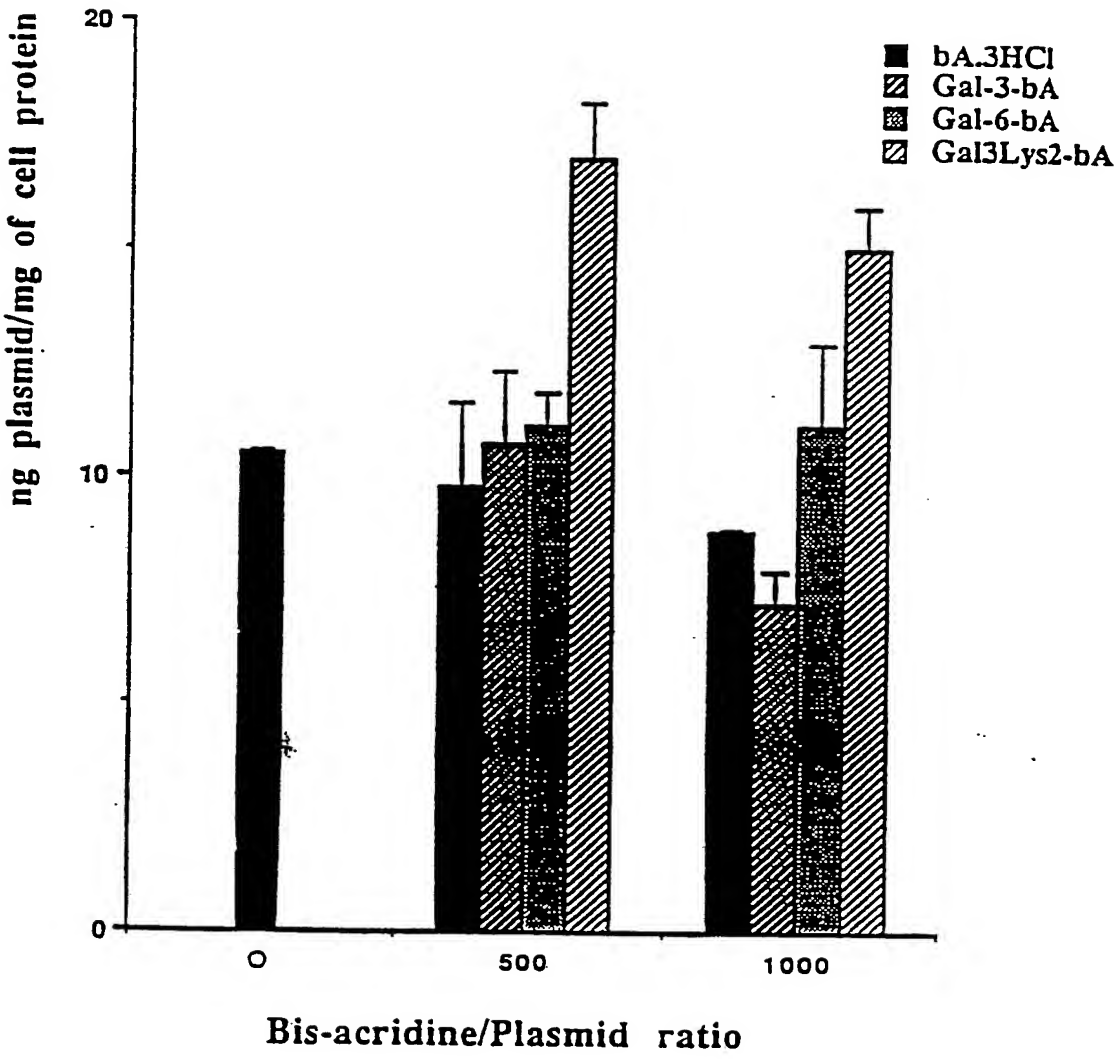

ng plasmid/mg of cell protein
20
10
0
bA.3HCl
Gal-3-bA
Gal-6-bA
Gal3Lys2-bA
0
500
1000
Bis-acridine/Plasmid ratio

FIGURE 12

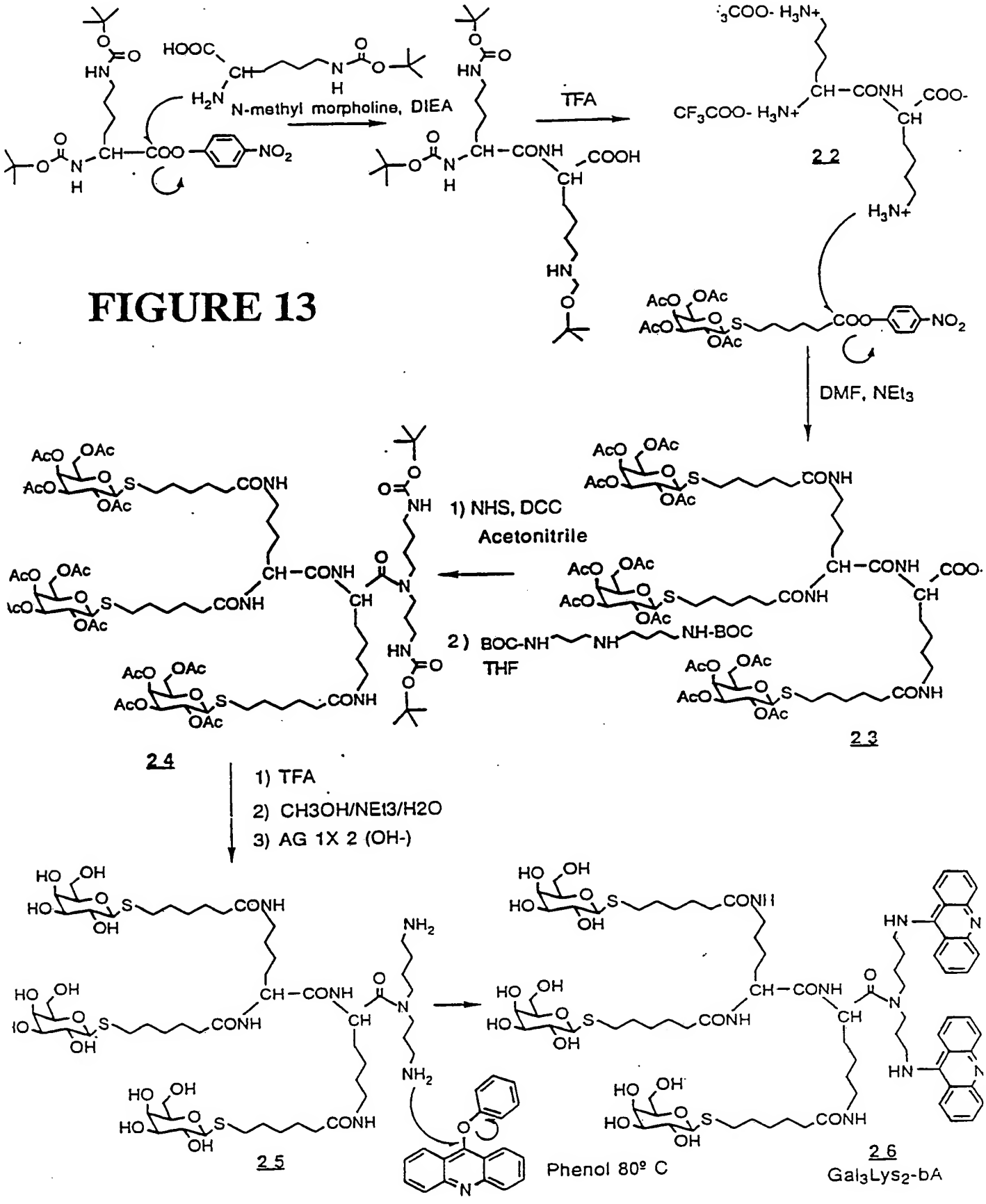

N-methyl morpholine, DIEA
TFA
2 2
DMF, NEt3
1) NHS, DCC
Acetonitrile
2) BOC-NH NH NH-BOC
THF
2 4
2 3
1) TFA
2) CH3OH/NEt3/H2O
3) AG 1X 2 (OH-)
2 5
Phenol 80° C
2 6
Gal3Lys2-bA

FIGURE 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/03406

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :Please See Extra Sheet.
US CL :435/172.3; 514/44; 536/23.1; 540/1; 544/1, 7; 546/1; 800/205

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/172.3; 514/44; 536/23.1; 540/1; 544/1, 7; 546/1; 800/205

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, Volume 112, issued 1990, G. Caminati et al., "Photophysical Investigation of Starburst Dendrimers and Their Interactions with Anionic and Cationic Surfactants", pages 8515-8522, see entire document.	1-20, 22-27, 29-32, 39-48
Y	PHARMACEUTICAL RESEARCH, Volume 9, number 10, issued 1992, J-Y Legendre et al., "Delivery of Plasmid DNA into Mammalian Cell Lines Using pH-Sensitive Liposomes: Comparison with Cationic Liposomes", pages 1235-1242, see entire article.	1-20, 22-27, 29-32, 39-48

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 14 JULY 1993	Date of mailing of the international search report 10 AUG 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer BRIAN R. STANTON Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/03406

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO, A, 91/04753 (BAER ET AL.) 18 APRIL 1991, see entire patent application.	1-20, 22-27, 29-32, 39-48
Y	WO, A, 90/09786 (SMITH ET AL.) 07 SEPTEMBER 1990, see entire application.	1-48
T	WO, A, 93/05162 (EPAND ET AL.) 18 MARCH 1993, see entire application.	1-48
T	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, USA, Volume 90, issued February 1993, Legendre et al., "Cyclic amphipathic peptide-DNA complexes mediate high-efficiency transfection of adherent mammalian cells", pages 893-897, see entire document.	1-20, 22-27, 29-32, 39-48
Y	BIOCHIMICA ET BIOPHYSICA ACTA, Volume 817, issued 1985, C.J. O'Conner et al., "Bile salt damage of egg phosphatidylcholine liposomes", pages 95-102, see entire document.	1-20, 22-27, 29-32, 39-48
Y	PLANT CELL TISSUE AND ORGAN CULTURE, Volume 22, issued 1990, Z. Zhu et al., "Transformation of tobacco protoplasts with DNA entrapped in pH-sensitive liposomes", pages 135-145, see entire document.	1-20, 22-27, 29-32, 39-48
Y	WO, A, 91/07947 (FREY ET AL.) 13 JUNE 1991, see entire patent application.	1-48
P,Y	WO, A, 92/22635 (WU ET AL.) 23 DECEMBER 1992, see entire patent application.	1-48
Y	WO, A, 87/02061 (PROTTER ET AL.) 09 APRIL 1987, see entire patent application.	1-48
T	BIOCONJUGATE CHEMISTRY, Volume 4, issued 1993, J. Haensler et al., "Synthesis and Characterization of a Trigalactosylated Bisacridine Compound To Target DNA to Hepatocytes", pages 85-93, see entire document.	1-48
P,Y	US, A, 5166,320, (WU ET AL.) 24 NOVEMBER 1992, see entire document.	1-48

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, USA, Volume 87, issued June 1990, M. Cotten et al., "Transferrin-polycation-mediated introduction of DNA into human leukemic cells: Stimulation by agents that affect the survival of transfected DNA or modulate transferrin receptor levels", pages 4033-4037, see entire document.	1-48
Y	HUMAN GENE THERAPY, Volume 2, issued 1991, F.D. Ledley, "Clinical Considerations in the Design of Protocols for Somatic Gene Therapy", pages 77-83, see entire document.	1-48
Y	WO, A, 91/15501 (ROSE ET AL.) 17 OCTOBER 1991, see entire patent application.	1-48

INTERNATIONAL SEARCH REPORT

...ternational application No.
PCT/US93/03406

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (5): A61K 37/00, 37/02, 37/20, 47/00, 47/06; C08H 61/00, 61/12; C08L 65/02; C12N 5/00, 5/06, 5/08, 5/16, 5/22, 15/00, 15/06, 15/07, 15/11

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN-file CA
Search terms: polynucleo?, DNA, nucleic acid?, oligonucleo?, phosphatidylethanolamine#, dileolyl, gramicidin, gramidicin S, tyrocidine, gala, peptide#, registry-2462-63-7, bile salt#, transfect?, deliver?, lipid?, liposome?, mask?, protect?, coat?, animal, vivo, degrad?, cation? dendrimer#, gene, therap?, intercalat?

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

I. Claims 1-20, 22-27, 29-32, 40 and 46-48, drawn to compositions for presenting polynucleotides to a cell, classified in Class 536, subclass 23.1 and the first appearing associated method of introducing polynucleotides into cells *in vitro*, classified in Class 435, subclass 172.3.

II. Claim 21, drawn to the first appearing organic compound, classified in Class 540, various subclasses depending upon the constituents.

III. Claim 28, drawn to the second appearing organic compound, classified in Class 540, various subclasses depending upon the constituents.

IV. Claims 33-35, drawn to the third appearing organic compound, classified in Class 540, various subclasses depending upon the constituents.

V. Claim 36, drawn to fourth appearing organic compound, classified in Class 540, various subclasses depending upon the constituents.

VI. Claims 37 and 38, drawn to the fifth appearing organic compound, classified in Class 540, various subclasses depending upon the constituents.

VII. Claims 39, 42 and 43, drawn to a method of preparing organic compounds classified in Class 540, various subclasses depending upon the constituents.

VIII. Claims 41-44, drawn to methods of introducing DNA into cell *in vivo*, classified in 514, subclass 44.

IX. Claim 45, drawn to gene therapy, classified in Class 514, subclass 44.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/03406

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
(Telephone Practice)
Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

THIS PAGE BLANK (USPTO)